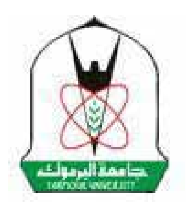
بسم الله الرحمن الرحيم

جامعة اليرموك

كلية التربية

قسم علم النفس والإرشاد التربوي



الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم في العلوم والرياضيات للصف العاشر الأساسي تبعاً لطريقة تقديم الاختبار

Differential item functioning of the national educational quality control test in mathematics and science for 10th grade according to mode of test presentation

إعداد

نجود علي فريحات

إشراف الأستاذ الدكتور

ساري سليم سواقد

حقل التخصص: القياس والتقويم التربوي

1435هـ-2014م

الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم في العلوم والرياضيات للصف العاشر الأساسي تبعاً لطريقة تقديم الاختبار

إعداد

نجود علي فريحات

بكالوريوس علوم أرض وبيئة، جامعة اليرموك 1996م

ماجستير قياس وتقويم تربوي، جامعة اليرموك 2007م

قدمت هذه الرسالة استكمالاً لمتطلبات الحصول على درجة دكتوراة الفلسفة في تخصص القياس والتقويم في جامعة اليرموك، إربد، الأردن.

أعضاء لجنة المناقشة

الأستاذ الدكتور ساري سليم سواقد .... رئيساً

أستاذ القياس والتقويم التربوي، جامعة مؤتة

الأستاذ الدكتور يوسف محمد السوالمة .... عضواً

أستاذ القياس والإحصاء التربوي، جامعة اليرموك.

الدكتور نضال كمال الشريفين .... عضواً

أستاذ مشارك في القياس والتقويم التربوي، جامعة اليرموك.

الدكتور زايد صالح بني عطا .... عضواً

أستاذ مشارك في القياس والتقويم التربوي، جامعة اليرموك.

الدكتور محمد منيزل العليمات .... عضواً

أستاذ مشارك في القياس والتقويم التربوي، جامعة آل البيت.

تاريخ مناقشة الأطروحة ٢٠١٤/١/٧

الإهداء

- إلى خالقي ومولاي الذي وفقني وفتح لي أبواب رحمته............... "الله" ربي لك شكر يليق بجلال وجهك وعظيم قدرتك
- إلى من ترفرف روحها حولي، وقلبها معي يخط على أوراقي إجابات تحكي عناءها من أجلي............... "أمي" لن أنساك، أزف إلى روحك بشرى نجاحي، وأدعو الله أن تكون الجنّة مثواك.
- إلى من جعله الله سبب وجودي، إلى من دعا لي في صلاته بين الركوع والسجود............... "أبي" أطال الله لي في عمرك وحشرك مع الأولياء والصالحين
- إلى من أتمنى لقاءه يوم الخلود وأتمنى أن يكون مكانه في الفردوس الأعلى............ أخي "وصفي" رحمك الله واسكن روحك في فسيح جنانه
- إلى رفيق دربي الذي تحمل من أجلي كل المصاعب, إلى من شدّ من أزري ووقف إلى جانبي بلا مقابل أو قيود.................. "زوجي الغالي محمد"
- إلى من أشتاق إليهم في سري وفي علني, إلى من تُهون ابتسامتهم علي مصائب الدنيا, فلذات كبدي الذين بعدت عنهم الساعات تلو الساعات............. أولادي "معتصم, يزيد, حمزة, وجنى"
- إلى إخوتي وأخواتي الذين زرعوا لي رمال الصحراء وروداً....... ومعهم أحس بجمال الوجود "فيصل, محمد, احمد, سلطان, ورود, بدر, دعاء, خالد"
- إلى جميع الأهل والأقارب, والأصدقاء والزملاء الطيبين

إليهم جميعًا أهدي هذا العمل المتواضع.

الباحثة

نجود فريحات

الشكر والتقدير

بسم الله الرحمن الرحيم

الحمد لله تعالى حمداً مباركًا يليق بجلال وجهه وعظيم رحمته, والشكر دومًا على نعمه وعطاياه, وأسأل الله جل جلاله أن يجعل عملي سبيلاً إلى رضاه, والصلاة والسلام على سيد المرسلين سيدنا محمد وعلى اله وصحبه أجمعين.وبعد.

يطيب لي بعد هذا الجهد المتواضع أن أتقدم بخالص الشكر وعظيم الامتنان إلى أستاذي الدكتور ساري سواقد المشرف على هذه الأطروحة, لما قدمه من وقت وجهد في سبيل المساعدة على اكتمالها بكل صدق وإخلاص, فجزاه الله عني كل الخير ومتعه بدوام الصحة والعافية.

كما أتقدم بالشكر والاحترام والتقدير إلى أساتذتي الأفاضل أعضاء لجنة المناقشة، الأستاذ الدكتور يوسف السّوالمة، والدكتور نضال الشريفين, والدكتور زايد بني عطا، والدكتور محمد عليمات, لتفضلهم بالموافقة على مناقشة هذه الرسالة, وللملاحظات القيمة التي أبدوها، والتي أسهمت في إثراء الدّراسة.

كما أقدم شكري وتقديري لأساتذتي الأفاضل في كلية التربية على ما قدموه لي من علم، ومعرفة خلال فترة الدراسة.

وأتقدم بالشكر الجزيل إلى المشرفتين التربويتين بسما, واسما المومني, والأستاذ احمد رفيق فريحات, والأستاذ علي العبدي والدكتور حسين الخطيب على ما قدموه لي من مساعدة في تدقيق وإخراج هذه الرسالة.

الباحثه

قائمة المحتويات

الموضوع	رقم الصفحة

الفصل الأول: خلفية الدراسة وأهميتها

الفصل الثاني: الدراسات السابقة

الفصل الثالث: الطريقة والإجراءات

الفصل الرابع: نتائج الدراسة

الفصل الخامس: مناقشة نتائج الدراسة

قائمة الجداول

قائمة الأشكال

قائمة الملاحق

الملخص

فريحات، نجود علي محمود. "الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم في العلوم والرياضيات للصف العاشر الأساسي تبعاً لطريقة تقديم الاختبار. "أطروحة دكتوراه، جامعة اليرموك، 2014. (المشرف: أ.د. ساري سليم سواقد).

هدفت هذه الدراسة إلى الكشف عن الأداء التفاضلي في اختباري العلوم والرياضيات من الاختبار الوطني لضبط نوعية التعليم وفقراتهما تبعًا لطريقة تقديم الاختبار (ورقي, محوسب).

ولتحقيق غرض الدراسة تم استخدام بيانات عينة مكونة من (2764) طالبًا وطالبة, من طلبة الصف العاشر الذين تقدموا للاختبارين للعام الدراسي 2011/2010، حيث بلغ عدد فقرات اختباري العلوم والرياضيات بصورتيهما النهائية من (25 ,24) فقرة على التوالي: من نوع الاختيار من متعدد، لكل فقرة منها أربعة بدائل، تقيس فقرات اختبار العلوم محاور تتعلق بالفيزياء، الكيمياء، علوم الأرض، والأحياء، في حين تتعلق فقرات اختبار الرياضيات بمحاور خاصة بالجبر، الهندسة والقياس, والإحصاء والاحتمالات.

تم تقدير معالم الفقرات باستخدام النموذج الثلاثي المعلمة، وتم الكشف عن الأداء التفاضلي باستخدام منهجية "الأداء التفاضلي للفقرة والاختبار"، حيث تم استخدام مؤشر "الأداء التفاضلي غير التعويضي (NCDIF)" كمؤشر للأداء التفاضلي للفقرة, ومؤشر "الأداء التفاضلي التعويضي (CDIF)" كمؤشر للأداء التفاضلي للفقرة من خلال ارتباطها بالدرجة الكلية على الاختبار. حللت بيانات الدراسة باستخدام البرامج الإحصائية التالية: (SPSS, BILOG-MG 3, Equate v2.1, Dfit v8.04) وكان من أبرز النتائج ما يلي:

- وجود أداء تفاضلي على مستوى فقرات اختبار الرياضيات تبعًا لشكل تقديم الاختبار بنسبة (8.33%) بواقع فقرتين من مجموع الفقرات التي يتكون منها الاختبار أظهرت الفقرة 28أداءً تفاضلياً منتظم لصالح الشكل المحوسب من الاختبار, في حين أظهرت الفقرة 13 أداءً تفاضلياً غير منتظم: تارةً لصالح الشكل الورقي, وتارةً أُخرى لصالح الشكل المحوسب من الاختبار.

- وجود أداء تفاضلي على مستوى فقرات اختبار العلوم تبعًا، لشكل تقديم الاختبار بنسبة (12%)، بواقع ثلاث فقرات من مجموع فقرات الاختبار جميعها، أبدت أداء تفاضليًا غير منتظم تارةً لصالح الشكل الورقي، وتارةً أُخرى لصالح الشكل المحوسب من الاختبار.

- عزت الدراسة أسباب ظهور الأداء التفاضلي على فقرات اختباري العلوم والرياضيات لصالح الاختبار المحوسب إلى انخفاض قيمة معامل الثبات لدرجات الاختبار على الشكل الورقي إذا ما قورن مع الشكل المحوسب، ويعود هذا إلى ارتفاع قيم معاملات الصعوبة للفقرات التي أبدت أداء تفاضليًا، والناتج عن أن طبيعة مضمون الفقرات، يحتاج إلى مهارات عقلية لا تتطلب من المفحوص وضع تدوينات خاصة على ورقة جانبية، للوصول للإجابة الصحيحة، فكان هناك سلاسة في التعامل مع الاختبار، باستفادة المفحوص من مزايا الاختبارات المحوسبة، مما رفع من شدة التركيز لدى المفحوص، وأبعده عن التخمين العشوائي، الذي لجأ إليه على الشكل الورقي للاختبار، في المقابل، الفقرة التي تطلبت مهارات عقلية، تستوجب وضع تدوينات خاصة بالمفحوص على ورقة جانبية، للوصول للإجابة الصحيحة، رفعت من شدة الإرباك لدى المفحوص على

الشكل المحوسب، وأعاقت من مدى استفادته من مزايا هذا الاختبار، فأصبح الشكل الورقي خيارًا مناسبًا يبعد المفحوص عن التخمين الذي يلجأ إليه على الشكل المحوسب.

- الفقرات التي أبدت أداءً تفاضليًا على مستوى الفقرة كانت قليلة، وبالتالي فإنه عمليًا لن يكون هناك ظلم كبير ما بين المفحوصين يعود إلى شكل تقديم الاختبار، وهو نتيجة مهمة لمطوري الاختبار، يمكن استخدامها بدقه لتقييم مدى امتلاك طلاب الصف العاشر لمهارات التعلم الأساسية في مجالي العلوم والرياضيات، واتخاذ قرارات مهمة عن واقع التعليم في الأردن.

- بشكل عام لم يبد الاختبار ككل عبر اختباري العلوم والرياضيات أداءً تفاضليًا دالاً إحصائيًا تبعًا لشكل تقديم الاختبار، وبالتالي قد لا يكون هنالك مشكلة في التوسع في تطبيق الاختبارات المحوسبة، كبديل صادق وعادل عن الاختبارات الورقية.

في ضوء ذلك، توصي الدراسة بإجراء المزيد من الدراسات لتحري الأداء التفاضلي على الاختبار الوطني لضبط نوعية التعليم تبعًا لشكل تقديمه، لتكون أكثر شمولية بأن تغطي المجالات الأربعة للاختبار (علوم، رياضيات، لغة عربي، لغة انجليزية)، وتتناول الصفوف الدراسية الرابع, الثامن، العاشر، والعمل على المقارنة بينها وفق نظرية الاستجابة للفقرة, كما تقدم الدراسة توصية لوزارة التربية والتعليم بالتوسع في تطبيق الامتحانات العامة حاسوبيًا.

الكلمات المفتاحية:

الأداء التفاضلي للفقرة، نظرية الاستجابة للفقرة، اختبارا العلوم والرياضيات من الاختبار الوطني لضبط نوعية التعليم، النموذج الثلاثي المعلمة.

الفصل الأول
خلفية الدراسة

يُعرف العصر الراهن بعصر الثورة العلمية والمعلوماتية والتكنولوجية؛ لهذا اهتمت النظم التربوية في مجتمع المعلومات، بإعداد الأفراد إعدادًا يؤهلهم للاستخدام الجيد للحاسبات وتكنولوجيا المعلومات. (الفار، 2000)

وقد شهد ميدان القياس والتقويم التربوي تطوراً كبيراً في استخدام الحاسبات، وتوظيف البرامج الحاسوبية، خاصة بعد ظهور نظرية الاستجابة للفقرة (**I**tem **R**esponse **T**heory, IRT) واستخدامها لتطوير الاختبارات، وما رافقها من استخدام البرامج الإحصائية الحاسوبية المختلفة؛ لمعالجة البيانات وتحليلها إحصائياً، وتقدير ما يعرف بمعلمة قدرات المفحوصين (Ability parameter)، ومعالم الفقرات (Items Parameter). وتمثل هذا التطور بالطريقة التي تقدم بها الاختبارات، من خلال تقديمها باستخدام الحاسوب (الاختبارات المحوسبة) (**C**omputerized **B**ased **T**est, CBT) بدلاً من تقديمها يدوياً (اختبارات الورقة والقلم) (**P**aper–**P**encil **T**est, PPT)، حيث أصبحت عملية تطبيق وتصحيح الاختبارات وتحليل نتائجها إحصائيًا باستخدام البرامج المحوسبة من الممارسات الشائعة في مجال التربية؛ لغايات اختيار الأفراد للوظائف المختلفة، أو تصنيفهم تبعًا لقدراتهم، أو تحصيلهم، أو الحصول على التراخيص، أو توجيه الطلاب مهنيًا، أو تحديد الصفوف المناسبة لهم. (Adarrage & Zacagnini, 1992; Westmeyer & Hagebock, 1992)

وقد تشهد السنوات القادمة مزيدًا من التوجه نحو استخدام الاختبارات المحوسبة كنهج سائد للاختبارات المدرسية، وربما يساعد على ذلك مدى توفر التكنولوجيا في المدارس، وذلك للمزايا العديدة التي تتمتع بها الاختبارات المحوسبة؛ فهي تحتاج إلى وقت أقل من حيث التطبيق واستخراج النتائج، وعرضها على شكل تقارير بالنسبة للطالب، والأطراف الأخرى المرتبطة به (مدير المدرسة، المعلم، ولي الأمر....الخ) من جهة، وهي أقل كلفة لأنها لا تحتاج إلى الورق والحبر لطباعتها، والأقلام للإجابة عنها من جهة أخرى. ويمكن أن تقدم تأثيرات ايجابية في المناهج التعليمية حيث تزود المتقدمين إلى الاختبار والمعلمين بتغذية راجعة تشخيصية فورية وشاملة، فأسئلة الاختبارات التي تتعلق بمهارة محددة، أو بمحتوى محدد من المعرفة يمكن تحديدها؛ لتشخيص جوانب القوة والضعف لديهم، مما يُمكن المعنيين من التخطيط السّليم، لتنمية جوانب القوة، ومعالجة جوانب الضعف (Becker, Frey & Pomplun, 2002; George, 2005; Jodoin, 2003; Wise & Plake, 1990).

وبما أن دخول هذا النوع من الاختبارات (المحوسبة) كطريقة لتقديم الاختبار يثير –كما هو الحال عند الإدخال لأي تغيير أو تطوير– العديد من التساؤلات، خاصة تلك التي تتعلق بقدرة الطلبة على التعامل معها، ومدى تأثير ذلك في نتائج الطلبة، مقارنة بالنتائج التي يحصلون عليها عند استخدام الأسلوب التقليدي المتبع لتقديم الاختبارات لهم (الاختبارات الورقية) من هنا تأتي هذه الدراسة لمحاولة فهم ما إذا كان شكل تقديم الاختبار (ورقي، محوسب) يؤدي إلى أداء تفاضلي على فقرات اختباري العلوم والرياضيات المستمدين من الاختبار الوطني لضبط نوعية التعليم، في محاولة لجعل مطوري الاختبارات يسعون إلى بناء وتطوير فقرات اختبارات

محوسبة عادلة وصادقة، تقيس الأداء الحقيقي المرتبط بالسمة المقيسة؛ للتوسع في تطبيق الاختبارات العامة حاسوبيًا كبديل للاختبارات الورقية.

مفهوم اختبار التحصيل (Achievement test)

تُعد اختبارات التحصيل وسيلة من الوسائل الهامة التي يعّول عليها قياس وتقويم أداء الفرد، بعد انتهائه من تعلم مادة دراسية معينة، أو موضوع دراسي معين، أو منهاج دراسي معين، وهي تقيس مقدار ما تحقق عند المتعلمين من النتاجات التعليمية، نتيجة مرورهم بخبرة تعليمية مقصودة وهادفة (علام، 2006)، فاختبار التحصيل "طريقة منظمة لتحديد مستوى تحصيل الطلبة لمعلومات ومهارات في مادة دراسية كان قد تم تعلمها مسبقاً بصفة رسمية، من خلال إجابته عن عينة من الأسئلة (الفقرات) التي تمثل محتوى المادة الدراسية". (الشريفين، 2009، ص.40)

وبناء على ما سبق فإن هناك تنوعًا في اختبارات التحصيل، واستخداماتها في المجالات التربوية، حيثُ يرى الين وين وجورج (Allen and Yen, 1979; George, 2005) أن وظيفة اختبارات التحصيل تتضمن: تشخيص نقاط القوة والضعف عند الطلبة، تقديم أدلة على تطور الطلبة، تقديم معلومات للتنبؤ بنجاح الطلبة في المستقبل، تقويم فعالية طرائق التدريس، كما ويمكن استخدامها لغايات الاختيار، والتصنيف، وتقويم البرامج التعليمية، والإرشاد.

ويتم تصنيف هذه الاختبارات بالاعتماد على معايير مختلفة من بينها الشكل الذي تقدم فيه للمفحوصين، حيث تصنف بالاعتماد على هذا المعيار إلى اختبارات الورقة والقلم والتي تمثل الشكل التقليدي الشائع لاختبارات التحصيل، حيث تقدم فقرات الاختبار بشكل ورقي، ويتم الإجابة عن هذه الفقرات بشكل يدوي باستخدام القلم، في المقابل هناك الاختبار المحوسب الذي

تقدم فيه فقرات الاختبار باستخدام أجهزة الحاسوب، ويتم الإجابة عن فقراته بشكل ألي، وأشار بويد (Boyd, 2003) إلى العديد من الاختبارات المحوسبة يمكن إيجازها كالتالي:

- الاختبار التقليدي المحوسب (**C**omputerized **B**ased **T**est,CBT).
- الاختبار المحوسب الذي يعتمد على برمجية تمكن من تغيير ترتيب الفقرات للطلبة عشوائياً.
- الاختبار التكيفي (Adaptive Testing) الذي يتلاءم مع القدرات المعرفية والمهارية للمتقدمين للاختبار.

ويمكن توضيح هذه الأنواع على النحو الآتي:

الاختبار التقليدي المحوسب:

هو باختصار النسخة الورقية المعتادة للاختبار، يتم فيه تطبيق الاختبار مباشرة على جهاز الحاسوب، ويتم تقديم جميع الفقرات لجميع المفحوصين بالترتيب نفسه. ولكن بدلاً من استخدام القلم والورقة يقوم الطلبة باستخدام الحاسوب.

الاختبار المحوسب الذي يعتمد على برمجية تمكن من تغيير ترتيب الفقرات للطلبة عشوائيًا:

يتم في هذا النوع من طرق تقديم الاختبار إعداد الاختبار، وكتابة فقراته على البرنامج المعدّ لذلك، وعند التطبيق، يقوم البرنامج بتقديم الاختبار نفسه للطلبة، والاختلاف فقط يكون في ترتيب الفقرات، وترتيب البدائل، حتى يحصل كل طالب على نموذج مختلف عن باقي الطلبة
الاختبارات التكيفية (Computerized Adaptive testing):

وهي اختبارات تعتمد على الحاسوب في اختيار الفقرات، فبعد أن يجيب الفرد عن فقرة اختبار إجابة صحيحة، تقدم له فقرة أكثر صعوبة تُختار طبقًا لصيغة رياضية معينة، وإذا أجاب عن الفقرة إجابة خاطئة، عندئذ تقدم له فقرة أقل صعوبة، ومن هنا جاءت تسمية "الاختبار التكيفي". ومع تطور الاختبارات التكيفية المحوسبة بدأ تطور الأساليب والإجراءات الخاصة بتقدير القدرة، وانتقاء الفقرات للفرد، من خلال الاختبار للوصول إلى دقة أكبر في تقدير القدرة وبأقل عدد ممكن من الفقرات.

مزايا ومحددات استخدام الحاسوب في الاختبارات وعلاقتها بالتحيز:

يمكن تلخيص مزايا استخدام الحاسوب في الاختبارات بالعمليات التالية: كتابة الأسئلة وبناء الاختبارات، تخزين الأسئلة بمواصفات محددة، تطبيق الاختبار، تصحيح الاختبار، تحليل الاختبار، تقرير يشمل نتائج الاختبار، كما تكمن أهمية الاختبارات المحوسبة في أمور تسهم في التخفيف على المعلم في الوقت والجهد الذي يبذله في الأعمال التعليمية الروتينية، واستثمار الوقت والجهد في تخطيط الدروس، وتمكن المعلم من تحليل نتائج الاختبار بطريقة سهلة. بالإضافة إلى أن الوسائل التعليمية الحديثة توفر بدائل ضرورية للطلبة ذوي الاحتياجات الخاصة، بحيث يمكن تغيير حجم الخط مثلا لذوي المشاكل البصرية، ويمكن باستخدام الاختبارات المحوسبة تغيير ترتيب الأسئلة لكل طالب، للحد من عمليات الغش (الصمادي، 2009).

وبما أن استخدام تكنولوجيا الحاسوب أصبح شائعًا في البيئات التعليمية والتقويمية، وانطلاقًا من حقيقة أنه عند تطبيق أي تكنولوجيا جديدة فإن الجهود تبذل في محاولة إثبات أن استخداماتها مفيدة لجميع عناصر العملية التعليمية، من خلال الحصول على علامات تفسر

بصورة صحيحة وصادقه بغض النظر عن الطريقة التي تقدم بها الاختبارات (ورقية أو محوسبة)، كأحد التطبيقات لهذه التكنولوجيا، فإن هذا يقود إلى التساؤل فيما إذا كانت هذه الاختبارات تتصف بالعدالة، وتعطي فرصاً متساوية لجميع المفحوصين، لإبداء المعرفة والمهارات المكتسبة ذات الصلة بغرض الاختبار (Roever, 2005)، فقد أشار عودة (2010)، إلى أن من الآثار السلبية البارزة للاختبارات، والتي حظيت باهتمام الباحثين في القياس النفسي والتربوي، التحيز بجميع أنواعه، مثل تحيز عينة الاسئله، تحيز صياغة الفقرات وخصائصها لطالب دون الأخر، والتحيز في محتوى الفقرات لصالح مجتمعات معينة، مما يعني تفاوت الطلاب في قدرات ليس لها علاقة بالبناء المقاس أو ما يسمى" Construct-irrelevance " كما أشار إلى ذلك كوبرين (kobrin, 2000) مثل:

- تدخل قدرة الطالب على استخدام الفأرة.
- تدخل قدرة الطالب على استخدام شريط التحويل عند قراءة النصوص الطويلة خلال الإجابة عن فقرات الاختبار، فالفقرات الطويلة التي تعرض على عدة شاشات في الشكل المحوسب، بالمقارنة مع نفس الفقرات التي تعرض على نفس الصفحة في الشكل الورقي للاختبار، تؤدي إلى وجود فروق في الأداء بين الطلبة على هذه الفقرات، وهذا سيؤثر في صدق علامات الاختبار، كما أن ذلك يسهم في تحديد سلبية أخرى للاختبار المحوسب، وهي عدم قدرته على تحقيق التكافؤ بين الشكل المحوسب والشكل الورقي للاختبار نفسه (Thompson, Thurlow & Moore, 2002).

كما أن هنالك متغيرات أخرى تُسهم في خلق عدم التكافؤ بين الشكل الورقي والشكل المحوسب للاختبار، وتعتبر من أوجه قصور الاختبارات المحوسبة، ومن هذه المتغيرات حسب

رأي كل من (Bracey, 1990; Davidson, 2003; Kobrin, 2000; Webster, & Compeau, 1996) ما يلي:

1- مهارات استخدام لوحة المفاتيح.

2- الفروق في هيكلة شكلي الاختبار.

3- الصور المختلفة للإخراج (مثل بنط الخط وحجمه، واللون، والوضوح).

4- الاختلاف في عدد الأسئلة التي تظهر على الورقة الواحدة عبر شكلي الاختبار.

5- ظهور رسائل تتعلق بحدوث خطأ ما أثناء الاختبار.

6- إعطاء معلومات عن عدد الأسئلة التي لم يتم الإجابة عنها.

7- عجز المتقدم إلى الاختبار في نظام الاختبار المحوسب عن مراجعة الاختبار، والرجوع إلى الخلف، وتغيير الإجابات.

وبالرغم من وجود هذه المظاهر للتحيز، إلا أنه لا غنى عن الاختبارات كأدوات للقياس في عملية التقويم؛ لذلك فإن التحقق من خلو فقرات الاختبارات المحوسبة من التحيز، يعد خطوة مهمة، كون التحيز أحد مهددات صدق علامات الاختبار، لذلك تزايد اهتمام مطوري الاختبارات ومستخدميها بقضية عدالة الاختبار–Test fairness– (Perrone, 2006)، حيث يعرف روفر الاختبار العادل بأنه الصادق نسبيًا لجميع المجموعات والأفراد، ويتيح للمفحوصين فرصًا متساوية لإظهار المهارات والمعارف التي اكتسبوها والتي تتصل بغرض الاختبار (Roever, 2005)، في ضوء ذلك، يتبع الباحثون منهجية للكشف عن التحيز، تتطلب إجراء مراجعة

وتحكيم شامله ودقيقة لجميع الفقرات، وذلك أثناء عملية تطوير الاختبار، وفي حال أن فقرة أو أكثر لم تنجح إجراءات التحكيم والمراجعة في اكتشافها، فإن الطرق الإحصائية تستخدم لاستكمال عملية الكشف عن الأداء التفاضلي للفقرات (Angoff, 1982; Berk, 1982; Sandifer, 2001, Uiterwijk & vallen, 2005).

وقد أعد هامبلتون وروجرز (Hambelton, & Rogers, 1995) قائمة مراجعة وتحكيم لكشف التحيز في محتوى الفقرات وبنيتها وصياغتها اللغوية، حيث يتطلب ذلك عرض هذه القائمة على فريق من المختصين في مجال القياس النفسي والتربوي، والمجال الذي يقيسه الاختبار، وممن لهم معرفة وخبرة في خصائص مجتمع الأفراد الذين صمم الاختبار ليطبق عليهم، بحيث يقوم هؤلاء بفحص الفقرات للكشف عن التحيز المحتمل فيها، بناءً على المحكات التي ترد في قائمة المراجعة، ليتم التعرف على العوامل التي تؤثر في أداء المفحوص على الفقرة، والمرتبطة بالعرق أو النوع أو المستوى الاقتصادي والاجتماعي، وذلك لإزالتها من الفقرات أثناء عملية تطوير الاختبار.

وفيما يتعلق بمفهوم تحيز الفقرة فقد تعددت وجهات نظر الباحثين والمختصين في المجالات النفسية والتربوية، فيرى هامبلتون وروجرز (Hambleton, & Rogers, 1995) أن الفقرة تعتبر متحيزة، إذا كان الفرق في المساحة بين منحنيات خصائصها (Item Characteristic Curve) في المجموعات المختلفة المتكافئة في القدرة ذا دلالة إحصائية، كأن يكونوا ذكورًا أو إناثًا، أو أن يكونوا من أعراق مختلفة، أي أن احتمال الاستجابة الصحيحة للفقرة مختلفة للأفراد في المجموعات الفرعية ممن لهم نفس مستوى القدرة.

كما يرى كاميلي وشيبارد (Camilli & Shepared, 1994) أن الفقرة تعتبر متحيزة، عندما تكون أكثر صعوبة لمجموعة دون أخرى، من نفس مستوى القدرة للسمة المراد قياسها في الاختبار.

وتعد الفقرة متحيزة، إذا أظهرت خصائص إحصائية متباينة للمجموعات المختلفة بعد مساواة قدرة المجموعتين على المقياس (De Ayala, Stapletion, & Dayton, 1999).

وعرف امبرتسون ورايز (Embretson, & Reise, 2000) الفقرة المتحيزة على أنها الفقرة التي تعمل بشكل مختلف لمجموعة ضد مجموعة أخرى.

وقد أورد بيرك (Berk, 1982) مجموعة من الإجراءات لتحرير فقرات الاختبار من التحيز هي:

- تطوير طرق وأساليب بناء اختبارات متحررة من التحيز، وتجنب الفقرات المتحيزة لمجموعة ما بسبب انتمائها إلى عرق أو جنس أو ثقافة معينة.
- مراجعة الاختبارات شائعة الاستخدام وتطويرها، عن طريق تقليل تحيزها، بتعديل أو استبعاد الفقرات المتحيزة.
- تطوير طرق وأساليب لكشف تحيز الاختبارات والفقرات في ضوء نتائج البحوث والدراسات.

التحيز والأداء التفاضلي للفقرة والاختبار (Bias and Differential Item and Test Functioning)

استخدم الباحثون في مجال القياس النفسي والتربوي في الأبحاث الحديثة مفهوم الأداء التفاضلي للفقرة (**D**ifferential **I**tem **F**unctioning, DIF)؛ للدلالة على تحيز الفقرة (De Gruijter & van Der kamp, 2005; Ellis & Raju , 2003; Zumbo, 2007)، والأداء التفاضلي للفقرة (DIF) هو مؤشر إحصائي للتعبير عن الفرق في احتمال الاستجابة الصحيحة للفقرة، بين مجموعات مختلفة، ولكنها متساوية في القدرة (Hambleton and Rogers, 1989; Pae, 2004; Swaminathan and Rogers, 1990)

وللتفريق بين الأداء التفاضلي للفقرة والتحيز فقد ذكر كاميلي وشيبرد (Camilli and Shepard, 1994)، أنه بالرغم من استخدام مفهوم الأداء التفاضلي للفقرة على أنه مرادف لتحيز الفقرة، إلا أن الأداء التفاضلي للفقرة يستخدم للكشف عن الفروق بين المجموعات متعددة الأبعاد، أو الاتجاهات لتحديد التحيز، وذلك عند تساوي الأفراد على البعد الأول للقدرة (Target Ability) في كلا المجموعتين، واختلافهم على البعد الثاني للقدرة (Nuisance Ability)، وهو البعد الذي لم يقصد قياسه في الفقرة، والذي يجعل احتمالية الإجابة الصحيحة عن فقرة في المجموعة الأولى تختلف عنها في المجموعة الثانية عند نفس المستوى من القدرة في البعد الأصلي.

وميز بينهما دورانز وهولند وايليس وراجو (Durance & Holland, 1993; Ellis & Raju, 2003) بأن الأداء التفاضلي للفقرة يهتم بالخصائص السيكومترية لها من حيث الكيفية التي تعمل بها الفقرة في المجموعات المختلفة، فيما يهتم التحيز بالمعنى الاجتماعي والإطار

النظري، حيث ينظر له على أنه مصطلح يتضمن معانيَ سلبية ترتبط بفكرة عدم العدالة، والاختلاف في إجراءات تطبيق الاختبار.

ويؤكد باي (Pae, 2004) على أن الأداء التفاضلي للفقرة يحدث عندما يكون هناك اختلاف في احتمال الإجابة الصحيحة عن فقرة ما بين أفراد مجموعتين من المفحوصين، من نفس المستوى في القدرة التي يقيسها الاختبار.

كما يذكر غروجتر وكامب (De Gruijter and Kamp, 2005) أنه إذا حدث، أو لوحظ الأداء التفاضلي للفقرة فمن المهم تحديد أسباب حدوثه، أي هل الفقرات تعتبر أقل أُلفه للمجموعة المستهدفة أكثر من المجموعة المرجعية؟ وإذا كان الأمر كذلك، هل فارقية الأُلفة غير متعلقة بالقدرة التي نحن مهتمون بها؟ فإذا كان الفرق بين أداء المجموعتين يعزى فعلا إلى عوامل ليس لها علاقة بالقدرة، نستطيع أن نستنتج أن الفقرة متحيزة، وهذا بالطبع أمر غير مرغوب به، فالتقليل من تحيز فقرات الاختبار يؤدي إلى التقليل من تحيز الاختبار ككل (Hambleton & Swaminathan & Rogers, 1991).

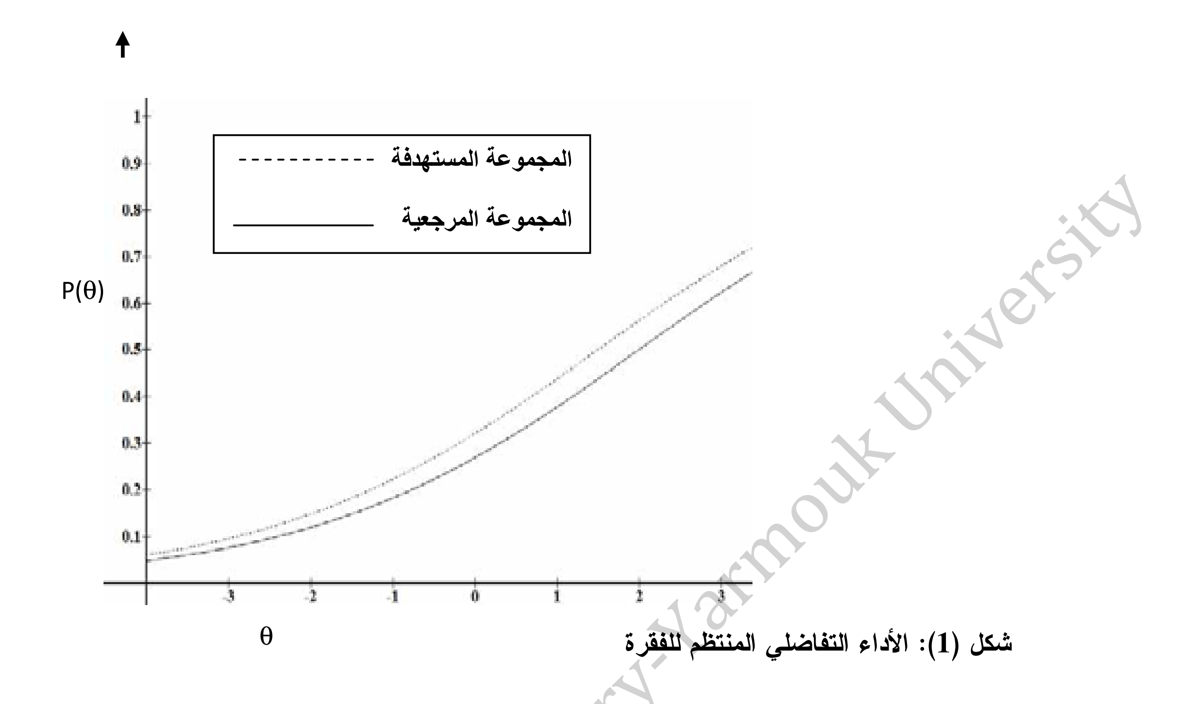
وللتفريق بين تحيز الفقرة (Item bias) وتأثير الفقرة (Item impact) فقد ذكر زامبو (Zumbo, 1999) أن تأثير الفقرة هو حصول المفحوصين من المجموعات المختلفة المتماثلة في مستوى القدرة على احتمالات مختلفة للإجابة عن الفقرة إجابة صحيحة، ويكون مرد هذا الاختلاف في الاحتمالات إلى وجود فروق حقيقية بين المجموعات في درجة امتلاك السمة التي تقيسها الفقرة، بينما يشير مفهوم تحيز الفقرة إلى أن مجموعة من المفحوصين لهم احتمالية أقل للإجابة عن الفقرة إجابة صحيحة من احتمالية مجموعة ثانية، وذلك بسبب وجود بعض الخصائص في فقرة الاختبار، أو ظروف التطبيق، والتي ليس لها صلة بغرض الاختبار.

وفي الوقت الذي تُميّز فيه أدبيات القياس بين الأداء التفاضلي والتحيز على مستوى الفقرة، فهي تميز أيضا بين الأداء التفاضلي والتحيز على مستوى الاختبار، فيذكر جنسن واوسترلند (Jensen, 1980; Osterlind, 1983) أن تحيز الاختبار حسب النظرية التقليدية هو خطأ منتظم في القياس، وهذا يجعل تأثيره في صدق الاختبار أكثر من تأثيره في ثباته، مما يؤدي إلى اختلاف دلالة العلامة الواحدة عند تقسيم الاختبار إلى مجموعات فرعية متكافئة، بينما يمثل الأداء التفاضلي للاختبار (**D**ifferential **T**est **F**unctioning, DTF) الفرق في الأداء على الاختبار بين المجموعات المرجعية والمستهدفة.

وتصنف نظرية الاستجابة للفقرة الأداء التفاضلي إلى نوعين تبعاً لطبيعة التفاعل بين مستوى القدرة، وعضوية المجموعة Hambleton & Rogers, 1989; Pae, 2004; (Swaminathan & Rogers, 1990):

الأداء التفاضلي المنتظم (Uniform-DIF): وفيه يكون احتمال الإجابة الصحيحة لمجموعة من المفحوصين أعلى أو أقل منه في المجموعة الأخرى عند جميع مستويات القدرة؛ أي لا يوجد تفاعل بين مستوى الأداء وعضوية المجموعة، ويكون الفرق دائما باتجاه واحد، والاختلاف بين منحنيات الاستجابة للفقرة في معلمة الصعوبة فقط.

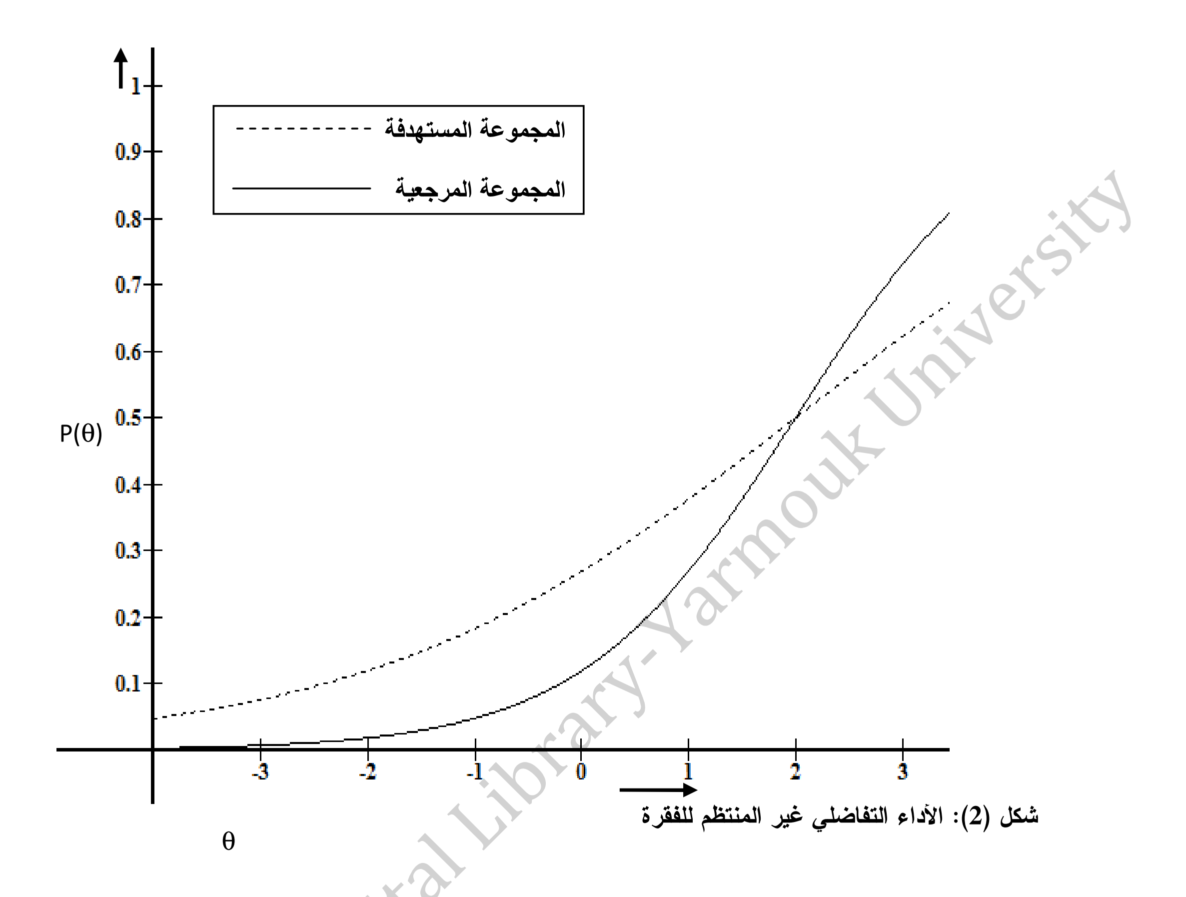
ففي الشكل (1): يلاحظ عدم وجود تفاعل بين مستوى القدرة وعضوية المجموعة، حيث يكون احتمال الإجابة الصحيحة عن الفقرة دائماً أكبر للمجموعة المستهدفة عند جميع مستويات القدرة.



شكل (1): الأداء التفاضلي المنتظم للفقرة

الأداء التفاضلي غير المنتظم (Non-Uniform DIF): في هذا النوع يوجد تفاعل بين مستوى الأداء وعضوية المجموعة؛ أي سيظهر الأداء التفاضلي لصالح المجموعة المستهدفة عند مستوى معين من القدرة، وعند مستوى قدرة أخر يظهر لصالح المجموعة المرجعية، ويكون الاختلاف بين منحنيات الاستجابة للفقرة في معلمة التمييز.

في الشكل (2): يلاحظ أن الأداء التفاضلي لصالح المجموعة المستهدفة عند المستويات الدنيا من القدرة، ولصالح المجموعة المرجعية عند المستويات العليا من القدرة. ويدل تقاطع المنحنيين على وجود تفاعل بين مستوى القدرة وعضوية المجموعة وبالتالي ظهور الأداء التفاضلي غير المنتظم.



شكل (2): الأداء التفاضلي غير المنتظم للفقرة

الأداء التفاضلي ونظرية الاستجابة للفقرة

مرت الاختبارات بمراحل متتابعة من التطور، وتعددت النظريات التي استندت إليها ومنها: النظرية الكلاسيكية التي سادت فترة من الزمن، وأُعتمدت مفاهيمها وافتراضاتها في بناء الاختبارات وتحليل نتائجها، ولكن تعاني هذه النظرية من بعض أوجه القصور التي حاول العلماء التغلب عليها، مما أدى إلى ظهور نظرية الاستجابة للفقرة، والتي يمكن باستخدامها مواجهة كثير من مشكلات القياس النظرية والتطبيقية التي عجزت عن مواجهتها النظرية الكلاسيكية (علام، 2006)، كما مهدت التطورات التي حصلت على نظرية الاستجابة للفقرة الطريق أمام تطبيقات جديدة في تحليل بيانات الاختبار وفقراته، مثل بناء الاختبارات Tests construction، معادلة الاختبارات Tests equating، بنوك الاسئلة Item banking،

الاختبارات التكيفية والمحوسبة Computerized and Adaptive testing، ودراسة الأداء التفاضلي لفقرات الاختبار. Differential test functioning (De Gruijter, & Van der kamp, 2005)

إن من أهم مميزات نظرية الاستجابة للفقرة على النظرية التقليدية في القياس، أن اقترانات خصائص الفقرة تمتاز بخاصية اللاتغاير(invariance) بين المجموعات الفرعية للمفحوصين على الاختبار، وذلك حسب النموذج المستخدم، بحيث يكون تقدير قدرة الفرد مستقلا عن عينة الفقرات التي تطبق عليه (Item Free)، وتكون فيها الخصائص السيكومترية للفقرات (مثل معاملات الصعوبة والتمييز) مستقلة عن عينة الأفراد المستخدمة في تقدير هذه الخصائص (Sample Free)، مما يجعل من نظرية الاستجابة للفقرة خيارًا ممتازًا للكشف عن الأداء التفاضلي للفقرات عبر المجموعات المستهدفة. (Hambleton & Swaminathan & Rogers, 1991)

تفترض نظرية الاستجابة للفقرة أنه يمكن التنبؤ بأداء الأفراد، أو يمكن تفسير أدائهم في اختبار نفسي أو تربوي معين بناء على خاصية أو خصائص معينة مميزة لهذا الأداء تسمى السمات (Traits)، ولكن من الصعب ملاحظة هذه السمات ملاحظة مباشرة، لذلك يجب تقديرها، أو الاستدلال عليها من أداء الأفراد الذي يمكن ملاحظته على مجموعة من فقرات الاختبار.(Hambleton and Swaminathan, 1985)

وتقوم هذه النظرية على بناء نماذج رياضية أو لوغاريتمية تمثل العلاقة بين قدرة المفحوصين (θ) وخصائص الفقرة من جهة، واحتمال حصول المفحوص على الإجابة الصحيحة

عن الفقرة من جهة أخرى، وقد وضع ماك دونالد (Mc Donald, 1982) تصنيفًا لهذه النماذج على النحو التالي:

1. النماذج أُحادية البعد (Unidimensional models)، مقابل متعددة الأبعاد (multidimensional models).
2. النماذج الخطية (linear models)، مقابل غير الخطية (Non–linear models).
3. النماذج ثنائية الاستجابة (dichotomous models)، مقابل متعددة الاستجابة (polytomous models).

ويشير هامبلتون وسوامنيثان وروجرز (Hambleton & Swaminathan & Rogers, 1991) إلى أن أكثر النماذج استخدامًا في تطبيقات نظرية استجابة الفقرة، هي النماذج اللوجستية، أحادية البعد، ثنائية الاستجابة. وبناءً على ذلك، فإن هنالك عددًا من النماذج والتي تعرف بنماذج السمات الكامنة (Latent Trait Models)، تسمح بالقياس الموضوعي، وتفسر استجابة الفرد للفقرة التي تنطوي على السمة التي تقيسها هذه الفقرة، والتي تصف منحنى خصائص الفقرة بالاعتماد على معالم الفقرة التي يشتمل عليها النموذج، كمتغيرات تحتاج للتقدير: (الصعوبة b_i، التمييز a_i، التخمين c_i)، وهناك ثلاث نماذج شائعة لمنحنى خصائص الفقرة، تقع ضمن النماذج أُحادية البعد التي تلاءم الفقرات ثنائية الاستجابة. (علام،2006؛ De Gruijter & Van Der Kamp, 2005):

1. النموذج اللوجستي أحادي المعلمة (الصعوبة b_i) One–parameter logistic model: يفترض أن معلمة صعوبة الفقرة هي المعلمة الوحيدة التي تؤثر في أداء المفحوص على فقرات الاختبار.

2. النموذج اللوجستي ثنائي المعلمة (الصعوبة b_i، التمييز a_i). Two–parameter logistic model يفترض أن معلمتي صعوبة الفقرة وتمييزها هما فقط المعلمتان اللتان تؤثران في أداء المفحوص على فقرات الاختبار.

3. النموذج اللوجستي ثلاثي المعلمة (الصعوبة b_i، التمييز a_i، التخمين c_i) Three–parameter logistic model: يفترض أن معالم الصعوبة والتمييز والتخمين للفقرة تؤثر في أداء المفحوص على فقرات الاختبار.

تستند هذه النماذج إلى افتراضات قوية يجب تحققها في البيانات، لكي تؤدي إلى نتائج يمكن الوثوق بها وهذه الافتراضات (Hambleton, Mazor and Jonse, 1993) هي:

- **أحادية البعد (Unidimensionality):** تفترض نماذج السمات الكامنة وجود قدرة واحدة تفسر أداء الفرد على فقرات الاختبار، وهذا الافتراض يصعب تحقيقه في كثير من الأحيان؛ لأن هناك عوامل أخرى ربما تؤثر في الأداء بالإضافة إلى قدرة الفرد، كالعوامل الشخصية والمعرفية، وظروف تطبيق الاختبار، فإذا لم تطابق البيانات النموذج بسبب انتهاك افتراض أحادية البعد، فإن إمكانية وجود الأداء التفاضلي يكون ممكناً (Hambleton & Swaminathan & Rogers, 1991).

- **الاستقلال الموضعي (Local Independence):** يقصد بهذا الافتراض أن تكون استجابات الفرد للفقرات المختلفة في الاختبار مستقلة استقلالا إحصائيا، وهذا يعني ألا تؤثر استجابة الفرد لإحدى الفقرات على استجاباته للفقرات الأخرى ويتضح هذا في: تحرر القياس من توزيع العينة المستخدمة (Sample Free).

تحرر القياس من مجموعة الفقرات المستخدمة (Item Free).

- **التحرر من السرعة (Speediness):** أي أن عامل السرعة لا يلعب دورا في الإجابة عن فقرات الاختبار.

- **منحنى خصائص الفقرة (Item Characteristic Curve, ICC):**

وهو عبارة عن دالة رياضية تربط بين احتمال الإجابة على فقرة إجابة صحيحة ويرمز لهذا الاحتمال P(θ) ، وبين قدرة المفحوص (θ) التي تقيسها مجموعة من الفقرات، أو اختبار معين.

يعتمد تعريف معامل الصعوبة Item difficulty **(b)** بالنسبة للفقرة **(i)** على طبيعة النموذج، أو عدد معالم النموذج، ففي النموذجين الأحادي والثنائي المعلمة تعرف صعوبة الفقرة بأنها: قيمة القدرة عندما يكون احتمال إجابة المفحوصين للفقرة إجابة صحيحة تساوي (0.5)، أما في النموذج اللوجستي ثلاثي المعلمة، فإن قيمة معامل الصعوبة للفقرة عبارة عن مقدار القدرة التي يكون عندها $Pi(\theta) = \frac{1+Ci}{2}$ حيث (C_i) هي قيمة معلمة التخمين، ويعتبر تمييز الفقرة Item discrimination **(a_i)** مؤشر يربط بين التغير في القدرة والتغير في احتمال الإجابة الصحيحة، وهي ليست الميل وإنما نسبة منه، ولكن كلما زاد الميل يزيد تمييز الفقرة، في حين تعبر معلمة التخمين"Pseud0-Guessing" parameter **(C_i)**: عن احتمال إجابة الفقرة إجابة صحيحة من الأفراد ذوي القدرة المتدنية (Hambeton & Swaminathan, 1985; Sheng, 2005). أما معلمة القدرة للمفحوص (θ) فهي تعبر عن السمة الكامنة التي تقيسها فقرات الاختبار، وقد تكون أي من: القدرات الإدراكية، أو التحصيل أو الكفايات الأساسية أو الخصائص الشخصية (Hambleton and Swaminathan, 1985).

وعند تناول الأداء التفاضلي كأحد التطبيقات المهمة لهذه النظرية، وجد أن المهتمين بقضايا القياس النفسي والتربوي قد توجّهوا إلى ابتكار طرق للكشف عن الفقرات التي تظهر أداءً تفاضليًا بين مجموعات المفحوصين قيد الاهتمام، . فظهرت طرق عديدة وأساليب إحصائية مختلفة، طبقت على جميع أشكال الاختبارات سواء ثنائية الاستجابة Dichotomous, أم متعددة الاستجابة Polytomous، تناولت مقاييس لخصائص نفسية مختلفة: معرفية وعقلية، بالإضافة إلى مقاييس الاتجاهات والاستعداد والتحصيل وغيرها، وهدفت تلك الطرق والأساليب الإحصائية؛ لتنقية الاختبارات من الفقرات التي تبدي أداء تفاضليًا؛ لتحقيق العدالة والمساواة بين المجموعات الفرعية المختلفة، بحيث تكون هذه الاختبارات صادقة لجميع الأفراد والجماعات، وتعطي فرصاً متساوية أمام المفحوصين، لإبداء المعرفة والمهارات ذات الصلة بغرض الاختبار.

ومن الطرق القائمة على نظرية الاستجابة للفقرة (Item Response Theory, IRT) والمستخدمة للكشف عن الأداء التفاضلي: طريقة نسبة الأرجحية العظمى (Likelihood Ratio Test)، والطرق القائمة على قياس المساحةArea Measures (Methods)، والطرق القائمة على منهجية "الأداء التفاضلي للفقرة والاختبار (**D**ifferential **F**unctioning of **I**tems and **T**ests Framework, DFIT)"

والطرق المتبعة في الكشف عن الأداء التفاضلي للفقرة وفق نماذج منحنى خصائص الفقرة، تعتمد على التشابه في منحنيات خصائص الفقرات بين المجموعات الفرعية لمجتمع إحصائي، بأن تكون الفقرات متساوية في صعوبتها وتمييزها وتخمينها لجميع المفحوصين، ممن

لهم نفس مستوى القدرة من المجموعات الفرعية المختلفة وذلك حسب النموذج المستخدم (Chung wang & Hui, 2004; Crocker & Algina, 1986)، إلا إذا كان:

$$P(Xi= 1/ T, G= 0) = P(Xi= 1/T,G = 1) \ldots\ldots\ldots\ldots(1)$$

حيث G=0 و G=1 تمثل المجموعات الأولى والثانية، و(T) تمثل السمة الكامنة أو(θ)، وتعني أن احتمالية الإجابة على الفقرة (i) بشكل صحيح عند نفس المستوى من القدرة في المجموعتين متساوية، ويعتبر انتهاك هذه المعادلة هو بمثابة تعريف لمفهوم الأداء التفاضلي للفقرة حسب نظرية استجابة الفقرة (Lord & Novick, 1968).

ويمثل النموذج اللوجستي ثلاثي المعلمة النموذج الشائع الاستخدام للكشف عن الأداء التفاضلي للفقرة؛ لأنه يضم المعالم الثلاث الممكنة للفقرة (**c_i, a_i, b_i**)، ويعبر عنه بالعلاقة العامة التالية التي تقيس احتمال إجابة المفحوص ذو القدرة (θ) على الفقرة **i** (Hambleton, Swaminthan & Rogers, 1991):

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-Da_i(\theta - b_i)}} - - - i = 1,2,\ldots,n \ldots\ldots(2)$$

حيثُ:

a_i: معلمة تمييز الفقرة.

b_i: معلمة صعوبة الفقرة.

C_i: معلمة التخمين للفقرة .

θ: معلمة القدرة.

e: الأساس اللوغاريتمي الطبيعي وتساوي(2.718).

D: عامل التدريج (Scaling Factor) وهو ثابت لجميع فقرات الاختبار ومقداره1.70

$P_i(\theta)$: احتمال أن يجيب المفحوص ذو القدرة θ عن الفقرة (i) إجابة صحيحة.

وقد صنف ايرنسون (Ironson) المشار إليه في (Hambeton & Swaminathan, 1985) طرق الكشف عن الأداء التفاضلي للفقرة وفق نظرية الاستجابة للفقرة إلى:

- الطرق التي تعتمد على مقارنة درجة مطابقة النماذج- نماذج (IRT)- للمجموعات (Comparison of the fit of the item response Models of the data), وهذه الطرق تقوم بمطابقة البيانات في كل مجموعة, ولكل فقرة مع نماذج نظرية استجابة الفقرة, وإذا اختلفت البيانات بين مجموعتين في مطابقتهما لنفس النموذج عند مستوى قدرة ما, عندئذ فإن الفقرة تبدي أداء تفاضليًا.

- الطرق التي تعتمد على مقارنة منحنى خصائص الفقرة (ICC) للمجموعات المتكافئة في القدرة (Comparison of Item Characteristic Curve)، وهذه الطرق تقوم بتقدير معالم الفقرات بين المجموعات, ورسم منحنى خصائص الفقرة لكل فقرة في كل مجموعة من المجموعات, فإذا كانت الفقرة متحررة من الأداء التفاضلي, فإن المساحة بين المنحنيات يجب ألا تتعدى معياراً معينًا وهو خطأ المعاينة.

- الطرق التي تعتمد على مقارنة معالم الفقرات بين المجموعات المتكافئة في القدرة (Comparison of Item Parameters).

في الطرق السابقة يتم تقدير معالم الفقرات والقدرة؛ ليتم استخدامها في رسم منحنيات خصائص الفقرة، لكن هذه الطرق تقوم بمقارنة هذه المعالم مباشرة بين المجموعات عن طريق وضع كل معلم في مصفوفة, وحساب إحصائي خاص يتبع توزيع مربع كآي، وبالرغم من

الاختلافات الإجرائية لهذه الطرق للتوصل إلى مؤشرات حول الأداء التفاضلي للفقرة, إلا أنها تتفق في اعتبار الاختلاف في منحنيات خصائص الفقرة (ICCs) للمجموعات الفرعية المتكافئة في القدرة, موضع الاهتمام مؤشرًا على الأداء التفاضلي لها، والتي تعتمد على قياس المساحة بين منحنيي خصائص الفقرة، وقد طور الباحثون عدة طرق لقياس المساحة بين المنحنيات للكشف عن الأداء التفاضلي للفقرة منها:

- طريقة الجذر التربيعِ لمربع الفرق بين منحنيي خصائص الفقرة, ويتم وفق هذه الطريقة تحديد المساحة بين المنحنيين عن طريق إحصائي خاص, هو الجذر التربيعِ لمربع الفرق بين منحنيي خصائص الفقرة في المجموعتين الفرعيتين المتكافئة في القدرة موضع الاهتمام (RMSD) **R**OOT **M**EAN **S**QUARE **D**EVIATION, حسب المعادلة (3):

$$RMSD = \sqrt{\frac{1}{600}\sum_{j=1}^{N}(P_{iA}(\theta_j) - P_{i_B}(\theta_j))^2} \dots\dots\dots (3)$$

حيثُ أن:

$P_{iA}\ \theta_j$: احتمالية الاستجابة الصحيحة للمجموعةA على الفقرة i عند مستوى القدرة θ_j.

$P_{iB}\ \theta_j$: احتمالية الاستجابة الصحيحة للمجموعةB على الفقرة i عند مستوى القدرة θ_j.

فإذا كانت قيمة الإحصائي (RMSD) أقل أو تساوي (0.05)؛ فإن ذلك يدل على عدم وجود دالة تفاضلية (Camilli & Shepard, 1994).

- طريقة هامبلتون وسوامنثان (Hambeton and Swaminathan, 1985) التي تتلخص إجراءاتها لحساب المساحة بين منحنيي خصائص الفقرة بالخطوات التالية:

1- اختيار النموذج المطابق للبيانات.

2- تقدير معالم كل فقرة في المجموعتين الفرعيتين المتكافئة في القدرة موضع الاهتمام كل على حدة.

3- معايرة (standardizing or calibration) معالم الفقرة في المجموعتين، لوضع جميع المعالم على نفس التدريج, ونلجأ إلى هذا الإجراء بسبب تقدير المعالم من مجموعات مختلفة، وإذا تمت معايرة معلمة القدرة فإن ذلك يتطلب عملية معادلة (Equating).

4- تقدير متصل القدرة من −3 إلى +3 في المجموعات المختلفة وتقسيم متصل القدرة إلى فئات صغيره بطول (0.005) فئة وعدد K من الفئات.

5- تحديد مركز كل فئة θ_k.

6- حساب قيم احتمالية الإجابة الصحيحة عن الفقرة في المجموعتين عند كل θ_k. ولكل الفئات التي عددها K.

7- إيجاد القيمة المطلقة للفروق بين احتمالات الاستجابة الصحيحة على الفقرة في المجموعتين عند كل θ_k، ثم تجمع هذه الفروق ولكل θ_k من مستويات القدرة المختلفة وبالرموز:

$$A = \sum |P_{i1}(\theta_k) \ - \ P_{i2}(\theta_k) \cdot \Delta_{\theta} | \ldots\ldots\ldots\ldots (4)$$

$P_{i1}(\theta_k)$: احتمالية الاستجابة الصحيحة للفقرة (i) في المجموعة الأولى عند مستوى القدرة (θ_k).

$P_{i2}(\theta_k)$: احتمالية الاستجابة الصحيحة للفقرة (i) في المجموعة الثانية عند مستوى القدرة (θ_k) .

فإذا تم الحكم على أن هذه المساحة صغيرة, فهذا مؤشر على أن الأداء التفاضلي للفقرة صغير أيضًا, أما إذا حكم على أن هذه المساحة كبيرة فهذا مؤشر على أن الأداء التفاضلي للفقرة كبير؛ أي أن الفقرة تبدي أداء تفاضليًا.

قام راجو (Raju, 1937–2005) خلال عمله في مجال القياس النفسي, بتطوير طريقتين للكشف عن الأداء التفاضلي: الطريقة الأولى (Raju's area measures) تقوم على حساب المساحة بين منحنيات خصائص الفقرة، والطريقة الثانية (الأداء التفاضلي للفقرة والاختبار، DFIT) تقوم على حساب الأداء التفاضلي للفقرة والاختبار ككل بالاعتماد على الطريقة المقترحة من رودنر(Rudner, Getson,& Knight, 1980a, 1980b)، بحيث توفر كلا الطريقتين منهجيه سهلة ومرنة، للكشف عن الأداء التفاضلي، باستخدام نظرية الاستجابة للفقرة.

الطريقة الأولى لراجو (Raju, 1988) عبارة عن طريقة يمكن من خلالها حساب المساحة بين منحنيات خصائص الفقرة ذات المعلمة الواحدة، أو المعلمتين، أو الثلاثة معالم، حيث يتم التعبير عن المساحة في النموذج الثلاثي المعلمة بالمعادلة التالية:

$$Area = (1-C)\left|\frac{2(a_2-a_1)}{Da_1a_2} in\left(1+e\frac{da_1a_2(b_2-b_1}{a_2-a_1}\right)-(b_2-b_2)\right| ....(5)$$

a_1: تمييز الفقرة للمجموعة الأولى. a_2: تمييز الفقرة للمجموعة الثانية.

b_2: صعوبة الفقرة للمجموعة الثانية. b_1: صعوبة الفقرة للمجموعة الأولى.

C: قيمة التخمين.

D: قيمه ثابتة وتساوي 1.7.

أما المساحة وفق النموذج الثنائي المعلمة فإنه يتم التعبير عنها وفق المعادلة التالية عندما تكون قيمة التخمين صفر أو تقترب منه:

$$Area = \left|\frac{2(a_2-a_1)}{Da_1a_2} In\left(1+e\frac{Da_1a_2(b_2-b_1}{a_2-a_1}\right)-(b_2-b_1)\right| ....(6)$$

وعند استخدام النموذج الأحادي المعلمة فان المساحة تكون عبارة عن القيمة المطلقة للفرق بين معلمتي الصعوبة للمجموعتين الفرعيتين المتكافئة في القدرة موضع الاهتمام. (Hambeton & Swaminathan & Rogers, 1991)

$$\textbf{Area} = |(b_2-b_1)| \textbf{ ....(7)}$$

كما أشار راجو (Raju, 1988, p.40) إلى أن "عدم تساوي معلمة التخمين بين المجموعتين قد تعطي فرقًا كبيرًا في المساحة بحيث تخرج عن مدى القدرة المحدد"؛ لذلك افترض أن الأخطاء المعيارية للمساحة تتوزع توزيعا طبيعيًا مع تثبيت معلمة التخمين في تعبيره الرياضي، وفي حال عدم تساوي معلمة التخمين للمجموعتين، فإنه عادة ما يتم تثبيت هذه المعلمة

عند(C=0.2) لحساب فرق المساحة بين منحنيات خصائص الفقرة Hambelton & Rogers, (Chung Wang & Hui Su, 2004; 1989; Pae, 2004)، وإيجاد نقطة قطع Cut-Off (Value) للمساحة الإحصائية التي تعتبر معيارًا للحكم على وجود الأداء التفاضلي للفقرة، أو عدم وجوده، وقد اقترح هامبلتون وروجرز (Hambelton and Rogers, 1995) طريقة لحساب هذه القيمة الحدية (نقطة القطع) على النحو الآتي:

1- تقسيم المجموعة المرجعية إلى مجموعتين عشوائيتين.

2- حساب معالم الفقرة والقدرة لكل مجموعة.

3- رسم منحنى خصائص الفقرة لكل مجموعة.

4- حساب المساحة بين منحنيات خصائص الفقرة في المجموعتين.

5- تكرار الخطوات السابقة عدة مرات وتثبيت أكبر فرق للمساحة كنقطة قطع أو معيار، فإذا كان الفرق بين المساحة للمنحنيين من المجموعتين العشوائيتين أكبر من المعيار، فإن ذلك يدل على أن الفقرة تبدي أداء تفاضليًا.

الطريقة الثانية لراجو(Raju, van der Linden, & Fleer, 1995) عبارة عن طريقة تستخدم للكشف عن الأداء التفاضلي للفقرة والاختبار (The Differential Functioning of Item and Tests, DFIT) باستخدام نظرية الاستجابة للفقرة، وهي طريقة تراعي عدم تثبيت معلمة التخمين أثناء الكشف عن الأداء التفاضلي، لما لها من أثر سلبي ومساس مباشر في دقة الكشف عنه، ومن المميزات التي تجعل من هذه المنهجية طريقة قوية ومرنة في الكشف عن الأداء التفاضلي ما يلي:

-يمكن تطبيقها على بيانات الاختبار سواء كانت ثنائية الاستجابة، أو متعددة الاستجابة.

-يمكن تطبيقها على النماذج أُحادية البعد، أو متعددة الأبعاد.

-تستخدم للكشف عن الأداء التفاضلي للفقرة والأداء التفاضلي للاختبار ككل.

-توفر ثلاث مؤشرات للأداء التفاضلي: اثنان على مستوى الفقرة وواحد على مستوى الاختبار يعبر عنه بمؤشر DTF، أما مؤشرات الأداء التفاضلي للفقرة فيعبر عنها من خلال مؤشر الأداء التفاضلي التعويضي (**C**ompensatory **D**if, CDIF)، ومؤشر الأداء التفاضلي غير التعويضي (**N**on–**C**ompensatory **D**if, NCDIF). (Raju, Oshima, Fortmann, Nering, & Kim, 2006)

-جاءت لمعالجة إحدى محددات طريقة حساب المساحة بين منحنيات خصائص الفقرة, التي تتمثل في أن الاختلافات ما بين منحنيات الفقرة عند جميع مستويات القدرة, تسهم بشكل متساوي في قياس الأداء التفاضلي للفقرة، حيث تتغلب هذه الطريقة على ذلك, بإيجاد قيمة التكامل بعد ضرب اقتران الكثافة للقدرة للمجموعة المستهدفة بمربع المساحة, لإعطاء وزن لكل قيمة من القيم كما يظهر في المعادلة التالية:

$$\text{NCDIF} = \int_{\theta} \mathrm{d}_i(\theta)^2 f(\theta) d\grave{e} \ldots\ldots\ldots (8)_$$

حيثُ $f()\theta$ اقتران الكثافة لــ (θ) في المجموعة المستهدفة.

إن استخدام هذه الطريقة للكشف عن الأداء التفاضلي يعتمد على مدى تحقق افتراضات نموذج الاستجابة للفقرة المستخدم لهذا الغرض، فبدون تحقق هذه الافتراضات تصبح هذه الطريقة غير مجدية، كما أن هذه الطريقة تعتمد على التقدير الدقيق لمعالم الفقرات، والتقدير

الدقيق لمعالم الفقرات يتطلب أحجام عينات كبيرة من المفحوصين، مما يعني أن طريقة(DFIT) لن تكون مناسبة للعينات الصغيرة، فحجم العينة المطلوب لكل مجموعة عند استخدام نماذج الاستجابة للفقرة ثنائية الاستجابة، يعتمد على عدد المعالم التي نرغب في تقديرها، فالنموذج الأحادي المعلمة يتطلب عينة حجمها (N>200)، أما النموذج الثنائي المعلمة فيحتاج إلى عينة حجمها (N>500)، في حين نحتاج إلى عينة حجمها (N>1000) لتطبيق النموذج الثلاثي المعلمة (Crocker & Algina, 1986).

خطوات استخدام طريقة (DFIT) في الكشف عن الأداء التفاضلي للفقرة:

1- معايرة الفقرات لتقدير معالم الفقرات وتقدير معلمة القدرة لمجموعتين من المفحوصين (المجموعة المستهدفة Focal group، والمجموعة المرجعية Reference group) كلٌ على حده، وذلك عن طريق استخدام برمجيات حاسوبية مثل: BILOG-MG 3 (Zimowski, Maraki, Mislevy & Bock, 2002).

2- وضع هذه المعالم على مقياس مشترك من خلال عملية المعادلة (equating process) باستخدام أي طريقه من طرق المعادلة (Kolen & Brennan, 2004), باستخدام برمجيات حاسوبية مثل: برنامج Equate (Baker, 1993) للنماذج ثنائية الاستجابة أو متعددة الاستجابة، أو برمجية (IPLINK) (Lee & Oshima, 1996)، أو برمجية (PIE) للنماذج ثنائية الاستجابة، أو برمجية (IRTEQ) (Han, 2007) للنماذج ثنائية الاستجابة أو متعددة الاستجابة.

3- حساب مؤشرات DIFT: التي تتضمن مؤشرات (الأداء التفاضلي التعويضي CDIF، الأداء التفاضلي غير التعويضي NCDIF، الأداء التفاضلي للاختبار DTF).

4- استخدام طريقة معلمة الفقرة التكرارية (**I**tem **P**arameter **R**eplication, IPR) لفحص الدلالة الإحصائية (Oshima, Raju & Nanda, 2006)

باستخدام مؤشر الأداء التفاضلي غير التعويضي للفقرة **N**on–**C**ompensatory **D**if, NCDIF للنماذج ثنائية الاستجابة، يتم حساب الفرق بين منحنيات خصائص الفقرة حيث يعرف هذا المؤشر بالاعتماد على راجو وآخرون (Raju et al, 1995) بما يلي:

$$NCDIF = E_F\,[d_i(\theta)^2] \ldots\ldots\ldots\ldots\ldots (9)$$

حيث (NCDIF): عبارة عن متوسط مربع المساحة ما بين منحنيي خصائص الفقرة للمجموعة المستهدفة والمجموعة المرجعية موضع الاهتمام.

ويعبر عن $d_i(\theta)$: بـ

$$d_i(\theta) = P_{iF}(\theta) - P_{iR}(\theta) \ldots\ldots\ldots\ldots\ldots (10)$$

حيث $d_i(\theta)$: عبارة عن الفرق بين احتمال الإجابة الصحيحة على الفقرة(**i**) عند مستوى قدرة معين ما بين المجموعة المرجعية والمجموعة المستهدفة قيد الدراسة.

$P_{iF}(\theta)$: احتمال الإجابة الصحيحة عند مستوى قدرة معين، باستخدام معلمة الفقرة المقدرة من المجموعة المستهدفة.

$P_{iR}(\theta)$: احتمال الإجابة الصحيحة عند نفس مستوى القدرة، باستخدام معلمة الفقرة المقدرة من المجموعة المرجعية.

وقد سمي هذا المؤشر بغير التعويضي؛ لأنه لا يأخذ بعين الاعتبار الأداء التفاضلي لباقي فقرات الاختبار, فهو يقيس الأداء التفاضلي لكل فقرة على حده على افتراض تحرر الفقرات الأخرى للاختبار من الأداء التفاضلي، وهو يفيد في تحري الأداء التفاضلي على مستوى الفقرة، وأيضاً تحديد شكل الأداء التفاضلي للفقرة فيما إذا كان منتظماً أو غير منتظم، لاتخاذ قرار حول تعديل الفقرة أو إزالتها.

أما على مستوى الاختبار (المكون من(n) من الفقرات) فإنه يتم التعبير عن الفرق بين العلامات الحقيقية (D) بما يلي:

$$D = T_F(\theta) - T_R(\theta) \quad \text{.................} (11)$$

$$D = \sum_{i=1}^{n} d_i \quad \text{.................}(12)$$

حيثُ يؤخذ الفرق D على مستوى درجات الاختبار, وقد عرف راجو وآخرون (Raju et al, 1995) الأداء التفاضلي على مستوى الاختبار (DTF) بأنه:

$$DTF = E_F[(D^2)] \quad \text{.................} (13)$$

حيث تؤخذ القيمة المتوقعة E_F عبر توزيع (θ) للمجموعة المستهدفة ولبيان العلاقة بين هذا المؤشر ومؤشر الأداء التفاضلي التعويضيCDIF فإنه يمكن كتابته بالصيغة التالية:

$$DTF = \sum_{i=1}^{n} CDIF_i \quad \text{................}(14)$$

أي أن مؤشر الأداء التفاضلي للاختبار عبارة عن المجموع التراكمي لقيم مؤشر الأداء التفاضلي التعويضي للفقرات التي يتكون منها الاختبار.

كما يتم التعبير عن مؤشر الأداء التفاضلي التعويضي CDIF لفقرة معينة من خلال:

$$CDIF_i = E_F(d_{iD}) = Cov(d_i, D) + \mu_{di}\,\mu_D \quad \text{.........................} (15)$$

حيثُ أن:

Cov: التغاير (التباين المشترك) ما بين d_i و D.

μ_D: الوسط الحسابي لــ D للمجموعة المستهدفة.

μ_{di}: الوسط الحسابي لــ (d_t) .

ويعتبر مؤشر الأداء التفاضلي التعويضي ((Compensatory DIF (CDIF) طريقة تأخذ بعين الاعتبار التباين المشترك للفقرة، عند قياس الأداء التفاضلي، عن طريق إيجاد علاقة بين الأداء التفاضلي للفقرة والأداء التفاضلي للاختبار, وهذه ميزة فريدة من نوعها تمكن من دراسة الأداء التفاضلي للفقرة، من خلال ارتباطها بالاختبار ككل؛ لمعرفة مدى إسهام كل فقرة في الأداء التفاضلي للاختبار, ومعرفة أثر حذف فقرة أو أكثر على الدلالة الإحصائية للاختبار، وبالتالي اتخاذ قرار حول الفقرات التي يجب إزالتها، ليصبح الاختبار خالياً من الأداء التفاضلي، وقد سمي هذا المؤشر بهذا الاسم؛ لأن قيم مؤشراتCDIF يمكن أن تكون ذات قيمة موجبة أو سالبة، وبالتالي تلغي الفقرات ذات قيمة CDIF الموجب الفقرات ذات CDIF السالب على مستوى الاختبار.

<u>معلمة الفقرة التكرارية (**I**tem **P**arameter **R**eplication, IPR):</u>

طريقة تستخدم لفحص الدلالة الإحصائية للمؤشر غير التعويضي للأداء التفاضلي، والتي تمثل طريقة لاشتقاق درجة قطع معينه لكل فقره من الفقرات (Oshima, Raju & Nanda, 2006)، من خلال عمل تقدير لمعالم الفقرة وحساب التباين والتباين المشترك لهذه التقديرات, وتكرار ذلك عدد كبير من المرات تصل في حدها الأقصى (1,000 تكرار للمجموعة المستهدفة)، بحيث يتم الحصول على عدد كبير من قيم NCDIFيتم ترتيبها وإعطائها رتب مئينية، تمثل توزيعًا تكراريًا لهذه القيم، ويتم تحديد درجة القطع عند قيمة NCDIF المرتبطة برتبة مئينيه من التوزيع

التكراري لقيم NCDIF، عند مستويات دلالة مختلفة تتراوح ما بين (0.001 إلى 0.5)، لتحديد فيما إذا كان مؤشر NCDIF دال إحصائيًا أم لا، وبذلك يتم الحصول على درجة قطع خاصة بكل فقرة، أما درجة القطع التي تفحص بها الدلالة الإحصائية على الاختبار فهي عبارة عن مجموع درجات القطع لقيم مؤشر NCDIF لجميع فقرات الاختبار.

مؤشر إزالة الفقرة (item removal index (2CDIF–NCDIF)):

مؤشر يستخدم لإزالة الفقرات من الاختبار ككل، بدءً بالفقرات التي يكون مؤشر الإزالة لها أعلى ما يمكن، والاستمرار في ذلك, حتى يصبح الاختبار غير دال إحصائيًا، بحيث توصف الفقرات التي يتم إزالتها، بأنها تبدي أداءً تفاضلياً، بالاعتماد على مؤشرات CDIF & ومؤشرات DTF).

الاختبار الوطني لضبط نوعية التعليم: (The National Education Quality Control Test)

تُعد مراقبة مستوى مخرجات النظام التربوي وتقييمه من المهام الضرورية لدفع عجلة التنمية في معظم المجتمعات الحديثة، حيث توفر نتائج الدراسات التقويمية الشاملة بيانات موضوعية تتعلق بمستوى إتقان الطلبة للكفايات الرئيسة, التي يسعى المجتمع إلى تنميتها، وصقلها لدى النشء، وتشكل بمجملها صورة المواطن الصالح وفق فلسفته، فمع انتشار حركات المساءلة, في ضوء مؤشرات الجودة ومعاييرها، وبهدف تقويم مخرجات العملية التعليمية, فقد اجتهدت وزارة التربية والتعليم، في تبني مشروع التطوير نحو اقتصاد المعرفة/المرحلة الثانية "THE **E**DUCATION **R**EFORM **F**OR **K**NOWLEDGE **E**CONOMY, ERFKEII"؛ لتطوير التعليم، عن طريق صقل المناهج، وتطوير أساليب التقويم ومصادر التعلم، بما في ذلك استخدام تكنولوجيا المعلومات والتعلم الالكتروني، والاستفادة منها في رفد العملية التعليمية،

وتطوير وتحسين أدوات القياس لمستويات الطلبة، بتنفيذ اختبارات وطنية تكشف عن مستوى جودة التعليم ونوعيته؛ ليصار إلى توظيف نتائجها في التخطيط لحركات التصحيح التربوي، وبرامج التطوير. وزارة التربية والتعليم (2009)

وفي هذا المجال قامت الوزارة منذ عام 2000م بإعداد مجموعة من الاختبارات منها الاختبار الوطني لضبط نوعية التعليم بشكليه: الورقي والمحوسب؛ لقياس مدى امتلاك الطلبة لمهارات التعلم الأساسية لديهم، ويهدف هذا المشروع إلى تحديد خط أساس مرجعي للأداء الوطني في مباحث محددة لصف محدد كل عام، كما يهدف إلى تحديد جوانب القوة في أداء كل طالب وجوانب الضعف، بهدف وضع حلول مناسبة لمعالجتها.

ويعد هذا الاختبار مرجعًا هامًا يمكن الاستفادة من نتائجه في إعداد الخطط العلاجية على مستوى المديريات والمدارس، واتخاذ القرارات التربوية حيثما يلزم ذلك، كما يعد وثيقة هامة يستفيد من نتائجها كل من: المعلمين, المدراء، أولياء الأمور، ومتخذي القرار في وزارة التربية والتعليم؛ سعيًا من الجميع إلى تحسين تعلم الطلبة.

ويجري الاختبار سنويا على كامل المجتمع لأحد الصفوف (رابع، ثامن، عاشر) في مباحث (اللغة العربية، اللغة الانجليزية، الرياضيات، العلوم), ويتم تزويد كل مدرسة من المدارس التي يطبق بها الاختبار بمستوى أداء كل طالب، مقارنه مع المستوى العام للمدرسة والمديرية والمملكة.

وفي ضوء ما يتم التوصل إليه من نتائج للاختبار، وتحقيقًا للأهداف المرجوة من الاختبار، فإن إدارة الامتحانات والاختبارات تصدر تقريرًا مفصلاً حول الاختبار، تُوصي بدراسته دراسة

متأنية ومتعمقة؛ لوضع الاستراتيجيات المناسبة، لمعالجة جوانب الضعف في أداء الطلبة، وتعزيز نقاط القوة لديهم. (وزارة التربية والتعليم، 2011)

مشكلة الدراسة

يشير الأدب والأبحاث المهتمة في استخدام الحاسوب في البيئات التعليمية والقياس والتقويم للعديد من القضايا المهمة التي تبحث في العلاقة بين الخبرة في مجال تكنولوجيا الحاسوب والمهارات الحاسوبية والاتجاهات نحوه, فبالرغم من الانتشار الواسع لاستخدام الحاسوب والانترنت, إلا أنه لا يزال هناك فروق في درجة ألفة الطلاب مع الحاسوب, لذلك فإن هنالك حاجة للوصول إلى مرحلة، يكون فيها لدى الطلاب المتساوين في القدرة نفس الفرصة في الأداء على فقرات الاختبار المحوسب، كما هو الحال في فقرات النسخة الورقية له، حتى لا يصبح هنالك مشكلة في التوسع في تطبيق الاختبارات المحوسبة كبديل عن الاختبارات الورقية، فبزيادة ألفة الطلاب باستخدام الحاسوب تصبح الفروق في الأداء على شكلي تقديم الاختبار ليس لها معنى، ولا تثير القلق والمخاوف في أن تكون فقرات الاختبار غير عادلة.

لذا فإن الهدف من هذه الدراسة هو تحري الأداء التفاضلي في اختباري العلوم والرياضيات من الاختبار الوطني لضبط نوعية التعليم وفقراتهما تبعًا لطريقة تقديمه (ورقي، محوسب)، المصمم لقياس مدى امتلاك طلبة الصف العاشر لمهارات التعلم الأساسية، وأنه من الأهمية بمكان التأكد من أن فقرات هذا الاختبار عادلة، وغير متحيزة لإحدى مجموعات المفحوصين على الأخرى على أساس طريقة تقديم الاختبار، باعتبار هذا الاختبار جهدًا وطنيًا لتفعيل دور التكنولوجيا في التقويم، مما يفسح المجال للتوسع في بناء وتطوير تطبيق الاختبارات المحوسبة لميزاتها المذكورة في الأدب النظري لهذه الدراسة.

كما جاءت هذه الدراسة لتستخدم طريقة "الأداء التفاضلي للفقرة والاختبار DFIT" للكشف عن الأداء التفاضلي على مستوى الفقرة والاختبار ككل, لضمان أن يكون الاختبار عادلاً وصادقًا، بخلو فقراته من الأداء التفاضلي، وإجراء تحرٍ شامل ودقيق، لمعرفة فيما إذا كانت الاختبارات المحوسبة بديلاً صادقًا وعادلاً عن الاختبارات الورقية.

وبشكل أدق فإن هذه الدراسة تحاول الإجابة عن الاسئلة التالية:

- هل تظهر فقرات اختبار الرياضيات الوطني لضبط نوعية التعليم لطلبة الصف العاشر (2011/2010) أداءً تفاضليًا ذا دلالة إحصائية عند مستوى الدلالة (α=0.05)، يعود إلى شكل تقديم الاختبار (ورقي مرجعي, محوسب مستهدف)؟
- هل تظهر فقرات اختبار العلوم الوطني لضبط نوعية التعليم لطلبة الصف العاشر (2011/2010) أداءً تفاضليًا ذا دلالة إحصائية عند مستوى الدلالة (α=0.05)، يعود لشكل تقديم الاختبار (ورقي مرجعي, محوسب مستهدف)؟

أهمية الدراسة

تُشكل الفقرة وحدة البناء الأساسية للاختبار، وعليه فمن الضروري تطوير فقرات اختبارات تتمتع بدرجة من اللاتغاير (الاستقرار) في تقدير معالمها عند تقديمها ورقيًا أو حاسوبيًا، للمجموعات المختلفة؛ حتى يكون هناك نوع من العدالة والمساواة عند تطبيق هذه الاختبارات، حيث تُعد خاصية خلو فقرات الاختبار من الأداء التفاضلي إحدى الخصائص المرغوب بها، وهذا سيساعد متخذي القرار على اتخاذ قرارات التقويم المناسبة، لمحاولة إحداث دمج وتكامل بين إجراءات القياس والتقويم، وعملية التعلم والتعليم.

وتكمن أهمية هذه الدراسة في سعيها لمعرفة فيما إذا كان لشكل تقديم الاختبار دور في ظهور الأداء التفاضلي في فقرات الاختبار الوطني لضبط نوعية التعليم – كون الأداء التفاضلي من المحددات الهامة لبناء وتطوير الاختبارات– مما سيساعد على محاولة معالجة ذلك من خلال توجيه الاهتمام لتدريب الطلبة على مهارات استخدام الحاسوب، وتعريضهم للعديد من الاختبارات التجريبية المصممة حاسوبيًا لإكسابهم الألفة والمهارة اللازمة لأن يكونوا مؤهلين لعصر الحوسبة وتكنولوجيا المعلومات، وتطوير برمجيات حاسوبية تقدم الاختبار بصورة لا تختلف عن الصورة الورقية، وبسرعة ودقة وسرية أكبر.

وبتضافر الجهود الوطنية ما بين وزارة التربية والتعليم، والجامعات، ومراكز الدراسات والأبحاث، بالتوسع بإجراء الدراسات القائمة على تحري الأداء التفاضلي للاختبارات العامة والوطنية، يتوقع أن تفيد نتائج هذه الدراسة في الإسهام في بناء وتطوير أدوات قياس وتقويم قادرة على إعطاء مؤشرات واستنتاجات صادقة، وغير متحيزة عن أداء الطلبة، مما سيساهم في اقتراح الأساليب والمشاريع التي تُحسن وضع الطلبة في اكتساب المهارات والمعارف العلمية الأساسية، وهذا سيساعد على زيادة المنافسة الإقليمية والعالمية، وتحسين نوعية المخرجات التعليمية، والذي يؤدي إلى تحسين نوعية التعليم العالي، وبالتالي رفد السوق المحلي والإقليمي والعالمي بطاقات قادرة على العمل بكفاءة وإتقان.

كما يمكن أن تفيد نتائج هذه الدراسة في زيادة دافعية المعلم، للمساهمة في تحسين وتطوير العملية التعليمية، باعتباره ركنًا مهمًا فيها، بتخليصه من مهمات استهلاك الوقت في إعداد وتصحيح الاختبارات، إذا ما تم الأخذ بنتائج الدراسة للمساهمة في تفعيل التوسع بتطبيق

الاختبارات حاسوبيًا، وإعداد بنوك للفقرات لبناء اختبارات الثانوية العامة، إذا كان هناك ثقة وعدم قلق من قبل الجميع بأن فقرات هذه الاختبارات ستكون عادلة وغير متحيزة .

وتستمد هذه الدراسة أهميتها من أنها تأتي لتستكمل ما توصلت إليه الدراسات السابقة وتستجيب لبعض توصيتها كدراسة ستونبيرغ (Stoneberg, 2004)، ودراسة بومريتش وبوردن (Pommerich & Burden, 2000)، فبرغم تزايد اهتمام متخصصي القياس النفسي والتربوي بموضوع الأداء التفاضلي لفقرات الاختبارات المحوسبة، إلا أن الجهود ما زالت قليلة على المستوى العالمي، وتكاد تكون معدومة على المستوى المحلي، مما يجعل من هذه الدراسة عنصر إسهام في هذه الجهود، كما توفر نتائج الدراسة معلومات كافيه لوزارة التربية تخدم الميدان في تطوير، وبناء اختبارات تخلو فقراتها من الأداء التفاضلي؛ حتى تكون عادلة في نتائجها.

متغيرات الدراسة

أولاً: المتغيرات المستقلة

متغير شكل تقديم الاختبار وله مستويان:

- الاختبار الورقي؛ ويتكون قبل المطابقة من (60) فقرة من نوع الاختيار من متعدد، على اختباري العلوم والرياضيات بواقع (30) فقرة على كل اختبار.
- الاختبار المحوسب؛ ويتكون قبل المطابقة من (60) فقرة من نوع الاختيار من متعدد، على اختباري العلوم والرياضيات بواقع (30) فقرة على كل اختبار.

- **ثانياً المتغيرات التابعة :**

- الأداء التفاضلي على الفقرات التي اكتشفتها منهجية "الأداء التفاضلي للفقرة والاختبار DIFT" من بين فقرات اختبار الرياضيات بأنها تبدي أداءً تفاضليًا تبعًا لشكل تقديم الاختبار.

- الأداء التفاضلي على الفقرات التي اكتشفتها منهجية" الأداء التفاضلي للفقرة والاختبار DIFT" من بين فقرات اختبار العلوم بأنها تبدي أداءً تفاضليًا تبعًا لشكل تقديم الاختبار.

التعريفات الاصطلاحية والإجرائية

منهجية "الأداء التفاضلي للفقرة والاختبار (DIFT)": الطريقة التي استخدمت في الدراسة للكشف عن الأداء التفاضلي للفقرة والاختبار من خلال مؤشرات خاصة بالأداء التفاضلي على الفقرة مثل (مؤشر الأداء التفاضلي غير التعويضي NCDIF, مؤشر الأداء التفاضلي التعويضي CDIF، ومؤشرات خاصة بالأداء التفاضلي للاختبار (DTF).

الأداء التفاضلي للفقرة (differential item functioning): هو مؤشر إحصائي للتعبير عن الفروق في الإجابة عن الفقرة بين مجموعتين من المفحوصين، تم اختيارهم من مجتمع الدراسة (طلاب الصف العاشر)، ممن هم في نفس المستوى من القدرة.

طريقة تقديم الاختبار (Test presentation): هي الطريقة التي تم فيها تقديم الاختبار الوطني لضبط نوعية التعليم للصف العاشر، وكيفية إجابة الطلبة عن فقراته، وذلك على النحو التالي:

- الطريقة الأولى (القلم والورقة)، وهي الطريقة التقليدية التي قام بها الطلبة بالإجابة عن فقرات الاختبار باستخدام القلم والورقة.
- الطريقة الثانية (الشكل المحوسب)، وهي الطريقة التي يقدم فيها نفس الاختبار التقليدي ولكن باستخدام الحاسوب، بحيث تتم الإجابة عن فقرات الاختبار من خلال الشبكة الالكترونية المحلية للمدارس (Etesting.Elearning.jo)

الأداء التفاضلي وفق طريقة تقديم الاختبار (differential item functioning according to mode of test presentation): دالة إحصائية تقاس بدلالة اختلاف الأداء على الفقرة لمفحوصين في نفس المستوى من القدرة، ولكنهم يختلفون تبعًا طريقة تقديم الاختبار (ورقي، محوسب).

الاختبار الوطني لضبط نوعية التعليم (the national education Quality_Control test): اختبار تُعده وزارة التربية والتعليم الأردنية/إدارة الاختبارات والامتحانات للعام الدراسي 2011/2010 لقياس مدى امتلاك الطلبة في الصف العاشر لمهارات التعلم الأساسية، في اختباري العلوم والرياضيات، بشكلين: ورقي ومحوسب.

الشكل الورقي للاختبار (Paper–Pencil Test, PPT): اختبار اختيار من متعدد رسمي ومكتوب ورقيًا، أعده فريق من المختصين في إدارة الاختبارات في وزارة التربية والتعليم في الأردن، يتم الإجابة عنه باستخدام الورقة والقلم بتظليل الدائرة التي تمثل رمز الإجابة الصحيحة، بقلم رصاص.

الشكل المحوسب للاختبار (Computerized Based Test, CBT): هو النسخة الورقية المعتادة للاختبار، تم تحويلها إلى اختبار معتمد على الحاسب الآلي، يتسم بمنهج ثابت، كما هو

الحال في النسخة الورقية الأصلية للاختبار. وبمعنى آخر فإن كل المتقدمين لأداء هذا الاختبار الحاسوبي يجيبون عن الأسئلة بالترتيب نفسه الذي تقدم فيه الأسئلة في النسخة الورقية.

اختبار العلوم (Science Test): اختبار اختيار من متعدد موزع على المحاور: الفيزياء، الكيمياء، الأحياء، علوم الأرض، بشكلين: ورقي ومحوسب، مستمد من الاختبار الوطني لضبط نوعية التعليم المكون من رزمتين اختباريتين: الأولى في (اللغة العربية والرياضيات)، والثانية في (اللغة الانجليزية والعلوم).

اختبار الرياضيات (Math Test): اختبار اختيار من متعدد موزع على المحاور: الجبر، الهندسة والقياس، الإحصاء والاحتمالات بشكلين: ورقي ومحوسب، مستمد من الاختبار الوطني لضبط نوعية التعليم المكون من رزمتين اختباريتين: الأولى في (اللغة العربية والرياضيات)، والثانية في (اللغة الانجليزية والعلوم).

المجموعة المستهدفة (Focal Group): وهي المجموعة التي يعتقد بأنها تتأثر بالأداء التفاضلي للفقرات، وتمثل الشكل المحوسب للاختبار الوطني لضبط نوعية التعليم للصف العاشر، على اختباري العلوم والرياضيات.

المجموعة المرجعية (Reference Group): وهي المجموعة المتبقية من أفراد المجتمع غير المنتمين للمجموعة المستهدفة، والتي تم مقارنة الأداء معها، وتمثل الشكل الورقي- الصورة التقليدية- للاختبار الوطني لضبط نوعية التعليم للصف العاشر كما هو متعارف عليه على اختباري العلوم والرياضيات.

افتراضات الدراسة

الاهتمام الذي توليه وزارة التربية والتعليم من خلال الحرص على بناء الاختبار بناءً صحيحًا، والتحقق من خصائصه السيكومترية قبل تقديمه للتطبيق، والحرص على تطبيق الاختبار في نفس الظروف، أو في ظروف متقاربة في جميع المدارس المطبقة للاختبار.

1- تكافؤ الطلبة الذين خضعوا للاختبار عبر طريقتي تقديمه، من حيث الاختيار العشوائي لهم للاستجابة لاختباري العلوم والرياضيات من الاختبار، وجدية الطلبة في الإجابة عن فقرات الاختبار في جميع المدارس المطبقة للاختبار بطريقتي تقديمه الورقية والمحوسبة.

محددات الدراسة

1- المشكلات الإدارية الخاصة بوزارة التربية والتعليم من خلال عدم الرغبة بإعطاء كافة البيانات المتعلقة بالاختبار على مجالاته الأربعة باعتبار أن هذه البيانات سرية، وأن بعض الوثائق المتعلقة به وثائق محمية، مثل جدول المواصفات.

2- اقتصرت الدّراسة على تحري الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم المقدم لطلبة الصف العاشر في جميع المدارس المطبقة للاختبار خلال العام الدراسي 2011/2010 ضمن بعد طريقة تقديم الاختبار (ورقي، محوسب).

3- اختيار اختباري العلوم والرياضيات من أصل رزمتين اختباريتين: الأولى في (اللغة العربية والرياضيات)، والثانية في (اللغة الانجليزية والعلوم)؛ للكشف عن الأداء التفاضلي في فقرات هذين الاختبارين، لتوفرها ورقيًا وحاسوبيًا

4- اقتصرت الدراسة على استخدام طريقة الارجحية العظمى الهامشية المستندة على مفاهيم نظرية الاستجابة للفقرة، لتقدير معلمة القدرة للمفحوصين، من أجل الحصول على نفس مستوى القدرة في مجموعات الدراسة.

5- اقتصرت الدراسة على استخدام منهجية راجو Raju:"الأداء التفاضلي للفقرة والاختبار (DIFT) "في النموذج اللوجستي الثلاثي المعلمة المستندة على مفاهيم نظرية الاستجابة للفقرة للكشف عن الأداء التفاضلي للفقرة، وما تم استخدامه من معالجات، وإجراءات إحصائية خاصة بالكشف عن الأداء التفاضلي للفقرة وفق هذه المنهجية، ومن برمجيات إحصائية متخصصة ومتوفرة.

الفصل الثاني

الدراسات السابقة

يشير عوده و ملكاوي (1992)، إلى أنه حين يختار الباحث مشكلة من مجال لم يكتب فيه إلا القليل، فإنه يواجه بمهمة تحديد الدراسات التي ترتبط بمشروع بحثه بطريقة غير مباشرة، وفي هذه الحالة يمكن أن يبحث عن دراسات سابقة، تضمنت عنصرًا أو أكثر من عناصر دراسته، ولكن في مجال أخر؛ لذلك وبعد مراجعة الأدب التربوي، والاطلاع على الدراسات التي اهتمت بالأداء التفاضلي حسب طريقة تقديم الاختبار: (ورقي محوسب)، لوحظ بأن معظم الدراسات التي تناولت الاختبارات المحوسبة والورقية كانت قائمة على المقارنة بين هذه الاختبارات، في ضوء عدة متغيرات، وأن الدراسات التي تناولت الأداء التفاضلي حسب متغير طريقة تقديم الاختبار كانت قليلة بالنسبة إلى دراسة الأداء التفاضلي بالنسبة إلى متغيرات أخرى، وربما لا يوجد أي دراسة بحثت الأداء التفاضلي لاختبارات وطنية تبعًا لطريقة تقديم الاختبار.

ولأغراض هذه الدراسة، تم الاطلاع على ما تيسر من الدراسات التي لها علاقة بموضوع الدراسة الحالية، وتم تصنيفها في أربع مجالات، من حيث مدى ارتباطها بموضوع الدراسة، وحسب الترتيب الزمني ضمن كل مجال، وتشمل هذه المجالات ما يلي:

1. دراسات تناولت الأداء التفاضلي لفقرات اختبارات وطنية ورقية ومحوسبة.
2. دراسات تتعلق بالمقارنة بين الاختبارات الورقية والمحوسبة وفق معايير مختلفة.
3. دراسات تناولت الأداء التفاضلي حسب شكل تقديم الاختبار (ورقي، محوسب).

أولاً: الدراسات التي تناولت الأداء التفاضلي لفقرات اختبارات وطنية:

قام هاوسر وكنغزبيري (Houser,& Kingsbury , 2004) بدراسة هدفت إلى معرفة إلى أي مدى تظهر اختبارات ايداهو (Idaho) الوطنية في التحصيل وفقراتها فروقًا جوهرية في الأداء عبر المفحوصين في المجموعات المختلفة، على أساس العرق والجنس. من خلال تطبيق هذه الاختبارات حاسوبيًا، حيث تكونت هذه الاختبارات، من فقرات اختيار من متعدد، قدمت للصفوف: الرابع، الثامن، العاشر في ثلاث مجالات هي: القراءة، مهارات اللغة، الرياضيات، بواقع (42) فقرة للصفين الرابع والثامن، على كل مجال من هذه المجالات، ومن (55 فقرة على مجال القراءة ,56 فقرة على مجال مهارات اللغة,60 فقرة على مجال الرياضيات) للصف العاشر. طبقت هذه الاختبارات على عينة تم اختيارها بشكل عشوائي من (24,170 طالبًا وطالبة من الصف الرابع، 24,530 طالبًا وطالبة من الصف الثامن، 22,569 طالبًا وطالبة من الصف العاشر)، تم استخدام نظرية استجابة الفقرة للكشف عن الأداء التفاضلي، باستخدام الاستراتيجية القائمة على عمل رايت ودوغلاس وليناكر ورايت ,Wright & Douglas, 1977 (Linacre & Wright, 1989) التي تسمى بطريقة ليناكر (Linacre), للكشف عن الأداء التفاضلي.

أظهرت النتائج أن فقرات اختبارات القراءة، اللغة، الرياضيات للصفين الرابع والعاشر لم تظهر أداءً تفاضليًا تبعًا لمتغير الجنس، بينما بلغت نسبة الفقرات التي أظهرت أداء تفاضليًا لمتغير الجنس على اختبار الرياضيات للصف العاشر 25% من الفقرات الكلية، ولتحديد المصادر المحتملة للأداء التفاضلي، تم فحص هذه الفقرات من قبل خبراء المحتوى، حيث تبين أن فقرة واحدة من الفقرات التي أظهرت أداء تفاضليًا كانت أسهل للذكور منه للإناث، كما تم إجراء فحص لمحتوى الاختبار، من حيث مدى مناسبته لجميع مستويات مهارات الطلبة التي

يجري اختبارها، حيث تبين أن الاختبار يركز في الغالب على المستوى العلوي من المهارات المتعلقة بأساسيات الرياضيات، وعلى بعض المهارات المتعلقة بالجبر، وأساسيات الهندسة، وأن (15%) من الطلاب كانت علاماتهم تقع على علامة النجاح أو أقل، وهي نسبة تفوق بمقدار ستة أضعاف النسبة المستخلصة من اختبار الصف الرابع، وتفوق بمقدار 1.7 مرة من النسبة المستخلصة من اختبار الصف الثامن، وعند فحص الموضوعات التي يتضمنها محتوى الفقرات، تبين أن (3) من الفقرات الخاصة بموضوعات الجبر، كانت أسهل للإناث منه للذكور، وأن اثنتين منهما كانتا أسهل للذكور منه للإناث.

وفي دراسة مشابهة على نفس الاختبارات، ولنفس الصفوف، أجراها ستونبيرغ (Stoneberg, 2004) هدفت إلى فحص الأداء التفاضلي لمتغيري: الجنس (ذكر، أنثى)، العرق (ابيض ، اسود)، على فقرات اختبار ايداهو (Idaho) الوطنية في التحصيل، من خلال تطبيق الاختبارات على عينة تم اختيارها بشكل عشوائي من (28,844 طالباً وطالبة من الصف الرابع، 28,756 طالباً وطالبة من الصف الثامن، 26,311 طالباً وطالبة من الصف العاشر). تم استخدام طريقتي (SIBTEST)، ومانتل هانزل لمربع كآي؛ للكشف عن الأداء التفاضلي، وتم استخدام إجراءات طريقة حجم الأثر؛ لمعرفة فيما إذا كانت الفروق دالة إحصائياً بشكل ذا معنى، أظهرت النتائج أن نسبة الفقرات التي أظهرت أداءً تفاضلياً تبعًا لمتغير الجنس تراوحت ما بين 7% على اختبار القراءة للصف الرابع، إلى 37%على اختبار الرياضيات للصف العاشر، وأن نسبة الفقرات التي أظهرت أداء تفاضليًا تبعًا لمتغير العرق تراوحت ما بين 7% على اختبار الرياضيات للصف الثامن، إلى 19% على اختبارات القراءة ومهارات اللغة للصف الرابع، وأوصت الدراسة بمراجعة الفقرات التي أظهرت أداءً تفاضلياً، من قبل خبراء المناهج قبل

استخدامها مرة أخرى، وإجراء المزيد من الدراسات المستقبلية، على الأداء التفاضلي لاختبارات اللغة.

أما دراسة الحمادنه (2006) فقد هدفت إلى مقارنة الأداء التفاضلي لفقرات اختبارات تحديد الكفاءة اللغوية في اللغة الانجليزية في الجامعات الأردنية (آل البيت، والعلوم والتكنولوجيا، واليرموك)، تبعاً لمتغيري: الجنس، فرع الثانوية العامة، وفق طريقتي مانتل – هانزل، وفرق المساحة بين منحنيات خصائص الفقرة، ومعرفة ما إذا كان الأداء التفاضلي يختلف باختلاف مستوى القدرة، ولتحقيق غرض الدراسة، تم اختيار ثلاثة نماذج من اختبارات تحديد الكفاءة اللغوية يتقدم لها المفحوصين الكترونياً في بداية كل عام جامعي، وكانت فقراتها من نوع الاختيار من متعدد، لكل فقرة أربعة بدائل، بلغ عدد فقراتها في الجامعات الثلاث (50، 50، 100) فقرة على الترتيب، تقيس في مجملها مهارات مشتركة في اللغة الانجليزية، طبقت على عينة تكونت من (1,935) مفحوص.

كان من أبرز النتائج: أكثر عدد من الفقرات التي أظهرت أداء تفاضليًا لمتغير الجنس كان في اختبار جامعة اليرموك، وبنسبة 24% تقريبًا من فقرات الاختبار. أكثر عدد من الفقرات التي أظهرت أداء تفاضليًا منتظمًا لمتغير فرع الثانوية العامة كان في اختبار جامعة العلوم والتكنولوجيا، بنسبة 21% تقريبًا من فقرات الاختبار لصالح الفرع العلمي وغير منتظمًا بنسبة 26% تقريبًا من فقرات الاختبار. اتفقت طريقة فرق المساحة مع طريقة مانتل – هانزل، على أن عددا من الفقرات تظهر أداء تفاضليًا منتظمًا لمتغيري الدراسة، وبنسب مختلفة، وكانت أعلى نسبة اتفاق في اختبار جامعة العلوم والتكنولوجيا بنسبة 40% تقريبًا تبعًا لمتغير فرع الثانوية العامة، كما أن طريقة مانتل – هانزل استطاعت أن تكشف عن أكبر عدد من الفقرات التي تظهر أداء تفاضليًا منتظمًا، فيما أظهرت طريقة فرق المساحة فاعليتها في الكشف عن الفقرات

التي تظهر أداء تفاضليًا غير منتظمًا. وقد عزت الدراسة أسباب ظهور الأداء التفاضلي في بعض الفقرات لصالح الإناث، إلى الاختلافات الثقافية والاختلاف في سياق الكلام، أو إلى محتوى وبنية ومدلولات الفقرات، بحيث تكون أسهل وذات أُلفة لدى الإناث دون الذكور، كما عزت الدراسة أسباب ظهور الأداء التفاضلي في بعض الفقرات لصالح الفرع العلمي إلى محتوى الفقرة أو بنيتها وصياغتها، أو لاحتوائها على مفردات شائعة ذات ألفة لدى الفرع العلمي بعيداً عن التخمين.

كما أجرى الريان (2010) دراسة هدفت إلى فحص دلالة الفروق بين متوسطات تحصيل طلبة الصف السادس الأساسي في الاختبار الوطني لمقرر الرياضيات، وفقًا لبعض المتغيرات، ولتحقيق ذلك تم تحليل علامات (3075) طالبًا وطالبة من طلبة الصف السادس الأساسي ممن تقدموا للاختبار الوطني لمقرر الرياضيات خلال العام الدراسي 2008/2007، موزعين على (94) شعبة دراسية في مدارس مديريات التربية والتعليم. أظهرت نتائج الدراسة وجود فروق ذات دلالة إحصائية بين متوسطات تحصيل طلبة الصف السادس الأساسي في الاختبار الوطني لمقرر الرياضيات وفقًا لمتغيرات: المديرية، جنس المدرسة، حجم الصف الدراسي، المؤهل العلمي للمعلم، وخبرته ومؤهله التربوي، في حين لم تكن الفروق دالة وفقًا لمتغيرات: جنس الطالب ،جنس المعلم، وتخصصه، كما تبين عدم وجود علاقة دالة إحصائيًا بين متوسطات تحصيل الطلبة وتقديرات معلميهم من قبل مدراء المدارس ومشرفيهم.

وفي دراسةٍ أجراها مبارك (2010) هدفت إلى الكشف عن الأداء التفاضلي في فقرات اختبار العلوم المكون من (97) فقرة في الدراسة الدولية بيزا (2006) كاختبار دولي طبق على المستوى الوطني ورقياً، تبعًا لمتغيرات: الدولة، اللغة، الانتماء لمنظمة التنمية والتعاون الاقتصادي، باستخدام طريقتي مانتل هانزل، ونسبة الارجحية، حيث طبق الاختبار على

(40,0000) مفحوص، تم اختيارهم بطريقة عشوائية، من مجتمع الدراسة، تم تحليل الاستجابات لعينة منهم مكونة من: (500 مفحوص من الأردن، 500مفحوص من الدول الأخرى)، (1000 لغتهم العربية، 1000لغتهم الانجليزية)، (2000من دول تنتمي للمنظمة، 2000 من دول لا تنتمي للمنظمة)، وقد كان أبرز نتائج الدراسة: وجود (63) فقرة ذات أداء تفاضلي لمتغير الدولة (الأردن مقابل غيرها) حسب طريقة مانتل هانزل، (90%) منها كان لصالح مجموعة الدول الأخرى. و(23) فقرة من أصل 46 فقرة حسب طريقة نسبة الارجحية، كان (70%) منها لصالح مجموعة الدول الأخرى. وأيضاً وجود (92) فقرة ذات أداء تفاضلي لمتغير اللغة (العربية مقابل الانجليزية) حسب طريقة مانتل هانزل، كانت جميعها لصالح اللغة الانجليزية، و(37) فقرة من أصل 46 فقرة حسب طريقة نسبة الارجحية، كانت جميعها لصالح اللغة الانجليزية. وجود (92) فقرة ذات أداء تفاضلي لمتغير الانتماء لمنظمة التنمية والتعاون الاقتصادي (تنتمي مقابل لا تنتمي)، حسب طريقة مانتل هانزل، كان (99%) منها لصالح مجموعة الدول المنتمية للمنظمة. و(18) فقرة من أصل 46 فقرة، حسب طريقة نسبة الارجحية كانت جميعها لصالح مجموعة الدول المنتمية للمنظمة.

تم تفسير النتائج المتعلقة بارتفاع نسبة الفقرات التي أظهرت أداء تفاضليًا لمتغيرات الدراسة، إلا أن الاختبار قد لا يكون متحررًا من المناهج، وأنه يتأثر بمناهج الدول المنتمية للمنظمة، بالإضافة إلى أن فقرات الاختبار الأصلية مكتوبة باللغة الانجليزية، وترجمتها قد يغير معناها، أو يؤدي إلى حذف معلومات مهمة، تجعل الإجابة عليها صعبة.

وتهدف دراسة كركايا (Karakaya, 2012) إلى تحديد الفقرات التي تظهر تحيزًا على مستوى الفقرة، تبعًا لمتغير الجنس، وذلك في الاختبارات الفرعية في العلوم والتكنولوجيا، والرياضيات التي قدمت للصفوف السادس، السابع، الثامن من عام 2009 ضمن اختبارات تحديد

المستوى (LDE)، المكونة من خمس اختبارات فرعية في اللغة التركية، الرياضيات، العلوم والتكنولوجيا، العلوم الاجتماعية، اللغات؛ وذلك لمعرفة مدى امتلاك الطلبة للمعارف والتحصيل العلمي، في البرامج التعليمية في كل مستوى من الصفوف الأساسية، لتستخدم الدرجات لغايات اختيار الطلبة للالتحاق بالمدارس الثانوية. وقد طبقت هذه الاختبارات على عينة تكونت من 22,624 طالباً وطالبة، تم اختيارهم بصورة عشوائية وزعت كالتالي: 69,13 من الصف السادس، 63,33 من الصف السابع، 93,74 من الصف الثامن، تم استخدام طريقة مانتل هانزل (MH)، لتحديد الأداء التفاضلي على فقرات الاختبار.

كشفت النتائج أن فقرات الاختبار الفرعي في العلوم والتكنولوجيا أظهرت أداء تفاضليًا بواقع فقرتين من الفقرات المقدمة للصف السادس، وثلاث فقرات من الفقرات المقدمة للصف الثامن، كانت ثلاثة منها لصالح الإناث، واثنتان لصالح الذكور، في حين أظهرت فقرات الاختبار الفرعي في الرياضيات أداء تفاضلياً تمثل بالفقرة الثالثة من الفقرات المقدمة للصف السادس، والفقرات المقدمة للصف السابع، والفقرتين الثانية والسادسة من الفقرات المقدمة للصف الثامن، منها ثلاث فقرات لصالح الذكور، وفقرة واحدة لصالح الإناث، وقد تم عرض هذه الفقرات على خبراء في القياس؛ لمعرفة فيما إذا كانت هذه الفقرات تظهر تحيزاً قائماً على الجنس، حيث أشار الخبراء إلى أنه لم تظهر أيًا من هذه الفقرات تحيزاً قائماً على الجنس.

ثانيًا: الدراسات التي تتعلق بالمقارنة بين الاختبارات الورقية والمحوسبة وفق معايير مختلفة:

قام جونسون وروس ووايز وبليك (Wise, Plake, Jonson, & Roos, 1992) بدراسة هدفت لبناء بنك فقرات يستخدم القياس التكيفي لاختبارات الاستعداد في القدرات المعرفية باستخدام النموذج الثلاثي المعلمة، قارنت الدراسة بين ثلاثة أنواع من الاختبارات: الاختبار التكيفي المحوسب، اختبار الورقة والقلم، الاختبار المحوسب العادي، حيث تم تحضير ثلاث

صور للاختبار من بنوك اسئلة برنامج كاليفورنيا للتقويم (CAP)، تم اختيار (55) فقرة لمستوى الصف الثالث لاختبار الورقة والقلم، و(62) فقرة لمستوى الصف السادس لاختبار الحاسوب، وتم تقدير القدرة باستخدام الأرجحية العظمى الهامشية، ثم اختيار الفقرات اللاحقة من التجمع بالاعتماد على القيم العظمى للمعلومات، ونهاية الاختبار كانت تعتمد إما على طول الاختبار (20) فقرة، أو أقل خطأ معياري=0,05 أيهما أسبق. تم تطبيق اختبارات الحاسوب، العادية والتكيفية بمعدل (30) طالب، تم اختيارهم بشكل عشوائي، بحيث خضعت كل مجموعة لنموذجين من الاختبارات، وكانت الأسئلة نفسها في النماذج الثلاثة للاختبارات: اختبار الورقة والقلم، الاختبار المحوسب العادي، الاختبار التكيفي المحوسب. أظهرت النتائج أن قيمة الانحراف المعياري منخفضة في كل من الاختبار المحوسب والاختبار التكيفي المحوسب، بالمقارنة مع اختبار الورقة والقلم، كما لوحظ اختزال في وقت الاختبار المحوسب (CT)، والتكيفي المحوسب (CAT). كما أظهرت الدراسة وجود ارتباطات ذات دلالة تتراوح بين 0,82 إلى 0,91 بين تقديرات القدرة، وبين رتب القدرة التي حسبت من الاختبارات الثلاثة.

أما دراسةُ فيسر (Viser, 1998) في جامعة جنوب إفريقيا، فقد هدفت إلى المقارنة بين كل من: الطريقة التقليدية (الورقة والقلم)، طريقة القياس التكيفي، طريقة الاختبار بواسطة الحاسوب لاختبارات القدرات المعرفية: (اللفظية وغير اللفظية) لطلبة المرحلة الثانوية في اختبار الاستعداد (G–SAT)، من حيث دقة القياس، عدد الفقرات المستخدمة، الوقت اللازم، حيث استخدم النموذج الثلاثي المعلمة في نظرية الاستجابة للفقرة، لمعايرة وتدريج الفقرات واستخراج معالمها، وتم استخدم برنامج محوسب في تطبيق الاختبارات بواسطة الحاسوب، وكانت فقرات الاختبار من نوع الاختيار من متعدد، بواقع (15) فقرة لكل من الاختبار اللفظي وغير اللفظي، والاختبارات المبرمجة، أشارت نتائج الدراسة إلى أن نتائج الطلبة كانت أفضل بواسطة الطريقة

التقليدية منها في القياس التكيفي أو الاختبارات المبرمجة، مع أن الخصائص السيكومترية (الثبات والصدق) متماثلة في الاختبارات الثلاثة، وعزت الباحثة السبب في ذلك، لتأثير استخدام الحاسوب على الطلبة، كما تبين أن القياس التكيفي يوفر 75% من الوقت مقارنة مع الاختبارين الأخريين.

وفي دراسة نوعية قام بها بوردن وبومريتش (Burden, Pommerich, 2000) هدفت إلى المقارنة بين استراتيجيات أخذ الاختبار، واستراتيجيات حل المشكلة، والانطباعات العامة عن الاختبار خلال تقديمه بشكلين: ورقي ومحوسب، وقد ركزت الدراسة على إجراء المقارنة على مستوى الفقرة، للتعرف على مميزات عرض الفقرات التي يمكن أن تسبب اختلافات على شكلي الاختبار. خضع (36) مفحوصًا (17) منهم من الذكور، و(19) من الإناث، على واحد من اختبارات الرياضيات، أو العلوم، أو القراءة، أو اللغة الانجليزية، تم توزيعهم حسب الرغبة على ثماني مجموعات، تبعًا لشكل تقديم الاختبار: (ورقي، محوسب)، ومجال محتواه (رياضيات، علوم، قراءة، لغة انجليزية)، كما تم إجراء أربع مقابلات واحدة لكل مجال من المجالات الأربعة للاختبار، وأيضاً تم عقد دورات، حول كيفية الإجابة عن الاختبارات المحوسبة والورقية، ويتكون اختبار الرياضيات من(15) فقرة اختيار من متعدد، يسمح للمفحوص باستخدام الآلة الحاسبة. تظهر الفقرات بشكل متتالي على النسخة الورقية، ويطلب من الطالب كتابة خطوات الحل للوصول إلى البديل الصحيح على ورقة الإجابة، في حين تظهر كل فقرة على شاشه واحدة على النسخة المحوسبة، ويطلب من الطالب كتابة خطوات الحل، للوصول إلى البديل الصحيح على ورقة بيضاء يتم تزويده بها لتحليلها لاحقًا. كما يتكون اختبار اللغة الانجليزية من جزئيين بواقع (15) فقرة على كل جزء، تتكون النسخة الورقية من ورقتين

تظهر القطعة والفقرات الخاصة بها على نفس الصفحة، في حين تعرض القطعة والفقرات الخاصة بها على نفس الشاشة عبر النسخة المحوسبة، أما اختبار القراءة فيتكون من قطعتين بواقع (15) فقرة على كل قطعة، تتكون النسخة الورقية من ورقتين تعرض القطعة والفقرات الخاصة بها على نفس الصفحة، في حين تشبه النسخة المحوسبة على الاختبار النسخة المحوسبة على اختبار اللغة الانجليزية. ويتكون اختبار العلوم من ثلاثة أجزاء تتراوح فقراتها ما بين (- 7 5) فقرات، تتكون النسخة الورقية من ورقتين، وتشبه النسخة المحوسبة على الاختبار النسخة المحوسبة على اختبار اللغة الانجليزية، باستثناء وجود رسومات وجداول ضمن فقرات الاختبار.

بالرغم من صغر عينة الدراسة إلا أن الملاحظات التي تم التوصل إليها من الدراسة، سلطت الضوء على بعض القضايا التي يجب أخذها بعين الاعتبار من قبل مطوري الاختبارات، لتحديد كيفية تقديم الاختبار، كما أن هناك العديد من العوامل التي تتفاعل مع قدرة المفحوصين للإجابة عن فقرات الاختبار، إضافةً إلى محتوى الفقرة ومن هذه العوامل: فواصل الصفحات والأسطر، ميزات تصميم القطع والفقرات، التظليل، خصائص الفقرة، بالإضافة إلى عوامل الحركة مثل: شريط التمرير، مراجعة الفقرات، معاينة الفقرات، قدرات الحذف للإجابات الخاطئة، كما تسهم خصائص المفحوصين بالعديد من الآثار الناتجة عن شكل تقديم الاختبار، وخاصة عدم جدية الممتحنين؛ لذلك ينبغي توخي الحذر، لضمان استجابة المفحوصين فقط لمحتوى الفقرة، وليس للميزات الكامنة المرتبطة بطريقة تقديم الاختبار، وتوصي الدراسة بإجراء المزيد من الدراسات في هذا المجال.

وقد هدفت دراسةُ جنفر وجون (Jennifer & John, 2003) إلى الكشف عن التكافؤ المعرفي لاختبارات القراءة باستخدام القلم والورقة وباستخدام الحاسوب. حيث تشير فرضية

الدراسة إلى أن المتقدمين للاختبار باستخدام الحاسوب، يكون لديهم تذكر أكبر للمعلومات الموجودة في النصوص، مما يكون له تأثير أكبر على أداءهم على الاختبار. تكونت عينة الدراسة من (48) طالباً وطالبة من جامعة عامة شمال شرق نيويورك، منهم (35) إناث، و(15) ذكور. تم تقسيم المشاركين عشوائياً إلى أربع مجموعات: (A, B, C, D) في كل مجموعة (12) طالباً. قدمت المجموعات (A, C) الاختبار باستخدام الحاسوب أولاً، ثم يليه الاختبار باستخدام القلم والورقة. والمجموعات (B, D) قدمت الاختبار باستخدام القلم والورقة أولاً، ثم يليه الاختبار باستخدام الحاسوب. وقد استخدمت لأغراض الدراسة فقرات من اختبار (GRE) الخاص بالقراءة والفهم، وأظهرت النتائج أن الفروق بين الطريقتين لم تكن دالة إحصائياً.

وفي دراسةٍ أُخرى أجريت من قبل بودمان وروبنسون (Bodmann and Robinson, 2004) هدفت إلى معرفة أثر تطبيق أشكال عديدة من الاختبار على درجات المفحوصين، وزمن إنهائهم للاختبار. تكونت الدراسة من تجربتين: التجربة الأولى هدفت إلى تحديد مدى الاختلاف في درجات المفحوصين وزمن إنهاء الاختبار على النسختين الورقية والالكترونية المكونتين من (30) فقرة اختيار من متعدد لكل منهما، على النسخة الالكترونية يتعرض المفحوص لفقرة واحدة في كل مرة، ويمكنه أن يراجع ويعدل الإجابة لمرة واحدة فقد دون التمكن من تأجيل أي فقرة لوقت لاحق، بالمقابل على النسخة الورقية تحتوي الصفحة الواحدة على حوالي ست فقرات بنفس الترتيب الموجود على النسخة الالكترونية. تكونت عينة الدراسة من (55) مفحوصاً في مساق علم النفس التربوي لهم خبرة سابقة في التعامل مع الحاسوب، من خلال خضوعهم لثلاث اختبارات الكترونية، دلت النتائج إلى أنه لا يوجد أثر لشكل تقديم الاختبار على درجات

المفحوصين، بينما كان هنالك أثر على زمن إنهاء الاختبار، حيث أنهى الطلاب الاختبار على النسخة الالكترونية بسرعة أكبر من زمن إنهاء الطلاب على النسخة الورقية. أما التجربة الثانية فقد هدفت إلى القضاء على بعض التباينات بين شكلي الاختبار من التجربة الأولى، الناتجة عن القدرة على تخطي أو مراجعة أو تغيير الإجابات، من خلال التركيز على أي اختلاف في الدرجات أو زمن إنهاء الاختبار، لذلك تم إزالة الاختبار الورقي المستخدم بالتجربة الأولى، واستعيض عنه بثلاث اختبارات الكترونية بمستويات مختلفة من المرونة:

- الاختبار الأول: يتعرض المفحوص لـــ (30) فقره مرة واحدة، يسمح له بالتنقل خلالها، من خلال شريط التمرير، لتخطي و مراجعة وتعديل الإجابات كلما أراد ذلك.
- الاختبار الثاني: يتعرض المفحوص لفقرة واحدة على كل شاشة، يسمح له خلالها بمراجعة وتغيير الإجابة، قبل تقديم كامل الاختبار.
- الاختبار الثالث مشابه جدا للنسخة الالكترونية في التجربة الأولى.

تم تطبيق هذه الاختبارات على عينة مكونة من (58) مفحوصًا تم توزيعهم عشوائيًا على واحدة من ثلاث مجموعات تجريبية، تم تحليل النتائج باستخدام تحليل التباين للقياسات المتكررة. دلت النتائج أنه لا يوجد أثر لشكل تقديم الاختبار على درجات المفحوصين، ولكن هناك أثر على زمن إنهاء الاختبار، حيث أن الاختلاف في درجة مرونة الاختبار تؤثر على زمن إنهاء الاختبار، حيث أكمل الطلاب الشكل الأقل مرونة بشكل أسرع من الشكلين الآخرين.

وفي دراسةٍ أجريت من قبل أيدن (AYDIN, 2006) بعنوان تأثير أجهزة الحاسوب على الاختبار وثبات الاتساق الداخلي لدرجات اختبارات (ESL) للمتعلمين، حيث تكونت عينة الدراسة

من (40) طالبًا وطالبةً في السنة الدراسية الثانية في قسم اللغة الانجليزية في جامعة اتاتورك، (20) منهم تقدم على النسخة الورقية، و(20) منهم تقدم على النسخة المحوسبة لاختبار اللغة الانجليزية، وكان عدد الطلاب وجنسهم من محددات الدراسة التي ارتبطت بقدرة مختبر حاسوب الكلية، وبتوزيع الجنس على مجتمع طلاب قسم اللغة الانجليزية، تم تحليل الدرجات باستخدام "كرونباخ الفا"، أظهرت النتائج أن الدرجات على الاختبار، وثبات الاتساق الداخلي لدرجات اختبار الكتابة المحوسب كانت دالة بشكل أكبر من تلك على الاختبارات الورقية.

وقام هوركاي وبينيت والين وكابلان ويان,Horkay, Bennett, Allen, Kaplan & Yan (2006) بدراسة هدفت إلى معرفة فيما إذا كان أداء المفحوصين على اختبار الكتابة عبر الانترنت وفقراته يختلف تبعًا لشكل تقديم الاختبار، وأثر ألفة الطلاب بالحاسب على أدائهم على الاختبار. ولتحقيق هدف الدراسة، تم اختيار(1,313) طالبًا وطالبة من طلاب الصف الثامن في المدارس الأساسية والثانوية في الولايات المتحدة ضمن برنامج التقويم الوطني للتقدم العلمي (NAEP) من عام (2002)، تم تقسيمهم إلى مجموعتين فرعيتين بحيث أعطي 715 منهم مهمة كتابة مقالتين باللغة الانجليزية باستخدام الورقة والقلم و أعطي 593 مهمة كتابة المقالتين باستخدام الحاسوب، بالإضافة إلى استجابة جميع المشاركين عن فقرات استبيان يتكون من (37) فقرة 10 منه معلومات عامة تتعلق بالجنس، العرق، ومستوى تعليم الوالدين، و21 فقرة تتعلق بخبرة المشاركين بالحاسوب، و6 فقرات منه تتعلق بمدى خبرة وتفاعل المشاركين مع الكتابة، وقد استخدم تحليل التباين للقياسات المتكررة ، تحليل التباين، وتحليل الانحدار لتحليل بيانات الدراسة. كشفت النتائج عدم وجود فرق ذو دلالة إحصائية ما بين أداء المشاركين على الاختبار المحوسب وأدائهم على الاختبار الورقي، كما أن مدى ألفة المشاركين بالحاسوب تستطيع أن تتنبأ

بشكل دال إحصائياً بأداء الطلاب على اختبارات الكتابة عبر الانترنت، وأن مدى اختلاف أداء المشاركين على شكلي تقديم الاختبار يعتمد على مدى ألفة الفرد باستخدام الحاسوب.

أما دراسة جونسون وغرين (Johnson & Green, 2006) فقد هدفت إلى استكشاف أثر شكل تقديم الاختبار (ورقي، محوسب) على لأداء على اختبار في الرياضيات، ومعرفة أهم العوامل التي تحفز الطلبة للعمل على الحواسيب في محاولة لفهم أسباب اختلاف أداء الأفراد حسب شكل تقديم الاختبار، ولتحقيق هدف الدراسة قُدّم الاختبار الوطني في الرياضيات في المملكة المتحدة لــ (104) من المشاركين من الفئة العمرية (10-11) سنة تم اختيارهم بناء على رغباتهم وموافقة الوالدين من أربعة من مدارس كامبريجشاير الابتدائية حيث تم توزيع عينة الدراسة على أربع مجموعات تجريبية متساوية بالقدرة بشكل عشوائي، وتم عقد دورة تدريبية للمشاركين حول كيفية استخدام الكمبيوتر تم خلالها خضوعهم لاختبارات تجريبية. يتكون الاختبار الوطني في الرياضيات من صورتين متكافئتين (A, B) كل صورة تحتوي على (10) اسئلة تغطي المستويات: الثالث، الرابع، والخامس من المنهاج الوطني في المملكة المتحدة بحيث أن كل صورة تتكون من سؤالين للمستوى الثالث وست اسئلة للمستوى الرابع وسؤالين للمستوى الخامس قدمت بشكلين ورقي والكتروني، طلب من المفحوصين إظهار خطوات حلهم للمسائل الرياضية عن الاختبار الالكتروني للوصول إلى الإجابة الصحيحة على ورقة بيضاء جانبية قدمت لهم، في حين ترك مكان مخصص لذلك بجانب كل فقرة من الاختبار على النسخة الورقية وذلك لتحليل هذه الخطوات التي يتبعها المفحوصين للوصول إلى الإجابة الصحيحة، و صمم الاختبار الالكتروني بحيث لا يستطيع المفحوص الانتقال إلى السؤال التالي إلا بالإجابة على السؤال الحالي، ولا يستطيع مراجعة وتعديل الاسئلة السابقة بينما كانت هذه القيود متاحة

للمفحوصين على النسخة الورقية. كما تم مراقبة سلوك المفحوصين أثناء الإجابة على اسئلة الاختبار، وإجراء مقابلات مع عينة فرعية من المفحوصين ، تم ترميز الأخطاء التي تم رصدها للطلاب خلال الاختبار بعد تصنيفها من قبل فريق العمل بالاختبار، وتم تحليل البيانات كميًا ونوعيًا.

أشارت النتائج عدم وجود فروق ذات دلالة إحصائية على الأداء العام على الاختبار تعود إلى شكل تقديم الاختبار ورقي، الكتروني، إلا أنه كان هناك فروق في الأداء على مستوى فقرات الاختبار مما يستدعي إجراء المزيد من الدراسات. التحليل الدقيق للبيانات أشار إلى أن نوع السؤال، طريقة طرح السؤال، والأرقام التي يحتويها قد تتفاعل مع طريقة عرض الاختبار في التأثير على مدى استعداد المفحوصين لإظهار خطوات حلهم والاستراتيجيات المتبعة في الحل، كما أشارت النتائج إلى أن أنواع معينة من الاسئلة في بعض المجالات قد يكون لها تأثيرات مختلفة حسب شكل تقديم الاختبار، وتوصي الدراسة بإجراء المزيد من الأبحاث حول إذا كان هناك أي ارتباط بين تفكير الأطفال وسلوكهم وطريقة تقيمهم من أجل دعم مدى صدق وموثوقية الاختبارات المحوسبة كبديل عن الاختبارات الورقية.

وهدفت دراسة أجراها بينيت وبراسويل وكابلان واورانج وساندن ويان (,Bennett) Braswell, Kaplan, Oranje, Sandene & Yan, 2008 إلى المقارنة بين الدرجات على اختبار الرياضيات بشكليه (الورقي، والمحوسب) .طبق الاختبار على عينة من طلاب الصف الثامن في المدارس الأساسية والثانوية في الولايات المتحدة والذين تم اختيارهم من برنامج التقويم الوطني للتقدم العلمي (NAEP) من عام (2001) بواقع (1,016) طالباً وطالبة استجابوا عن الاختبار المحوسب، و(954) طالباً وطالبة استجابوا عن الاختبار الورقي، تكونت أداة

الدراسة من اختبار في الرياضيات قدم بشكلين (ورقي، محوسب) بحيث تكون من 26 فقرة منها 16 فقرة اختيار من متعدد، و8 فقرات ذات إجابة قصيرة محدده، وفقرتين اسئله مقالية ذات إجابة محدده سمح للمشاركين باستخدام الآلة الحاسبة، كما استجاب المشاركين على استبيان مكون من 30 فقرة تتعلق بمعلومات ديموغرافية، ومعلومات حول خبرة المشاركين باستخدام الحاسوب، بالإضافة إلى خضوع المشاركين إلى دورة تدريبية حول كيفية استخدام الحاسوب والانترنت، تعرض المشاركين خلالها لاختبار تدريبي مكون من 20 فقرة اختيار من متعدد استخدمت لأغراض إجراء تحليل تباين مصاحب، كمتغير في الدراسة. وقد استخدم تحليل التباين المصاحب والنموذج الثلاثي المعلمة وفق نظرية الاستجابة للفقرة، لتحليل نتائج الدراسة، حيث أظهرت النتائج أن الاختبار المقدم بشكل محوسب كان أصعب بشكل دال إحصائيًا من الاختبار الورقي، كما أن مدى ألفة المشاركين بالحاسوب تستطيع أن تتنبأ بشكل دال إحصائيًا بأداء الطلاب على اختبار الرياضيات عبر الانترنت.

كما هدفت دراسة المومني (2009) إلى بيان أثر الطريقة التي تقدم بها نصوص الاستماع في اختبار مهارة الاستماع في اللغة الإنجليزية على الخصائص السيكومترية للاختبار وفقراته. ولتحقيق هدف الدراسة، تم تطبيق اختبار مؤشرات الأداء الخاص بالصف العاشر الأساسي والمكون من 30 فقرة من نوع الاختيار من متعدد، حيث قدمت نصوص الاختبار في ثلاث طرق مختلفة هي النصوص مسجلة بصوت أشخاص بريطانيين يتحدثون اللغة الانجليزية كلغة أم، النصوص مسجلة بصوت أشخاص أردنيين يتحدثون اللغة الانجليزية كلغة ثانية والنصوص مقدمة بصوت معلم الصف نفسه. طبقت على عينة مكونة من 1493 طالب وطالبة من طلبة الصف الأول الثانوي في المدارس الحكومية التابعة لمحافظة عجلون يدرسون في 59 شعبة

مختارة بالطريقة العشوائية حيث تم توزيع هذه العينة إلى ثلاث مجموعات تبعًا للطريقة التي قدمت فيها نصوص الاستماع. وأشارت نتائج الدراسة إلى وجود فروق ذات دلالة إحصائية في معاملات الثبات بين الطرق الثلاث للاختبار، وكانت هذه الفروق لصالح الطريقة الثانية، ووجود فروق ذات دلالة إحصائية في المتوسطات الحسابية لأداء الطلبة لصالح الطريقة الثالثة تعزى للطريقة التي قدمت فيها نصوص الاستماع.

وأجرى هغنز وباترسون وبوزمان وكاتز (Higgins, Patterson, Bozman & Katz, 2010) دراسة هدفت إلى معرفة مدى جدوى الانتقال من تطبيق اختبارات GED بصورتها الورقية إلى تطبيقها باستخدام الحاسوب وتأثير ذلك على درجات الاختبار، وخبرة المفحوصين، ولتحقيق هدف الدراسة طبقت اختبارات GED على مجال الرياضيات من أصل خمس مجالات تتكون منها الاختبارات بصورتين ورقية، ومحوسبة كل صورة مكونة من (25) فقرة خاصة بمهارات عملية بالرياضيات 20 منها اختيار من متعدد و5 فقرات اكمال الفراغ، سمح للمفحوصين باستخدام الآلة الحاسبة على أول13 فقرة منها بينما لم يسمح لهم بذلك على باقي الفقرات كما تم إجراء دراسة مسحية ورقية من ثلاثة أجزاء: فقرات تتعلق بمعلومات ديموغرافية، فقرات تتعلق بخلفية معرفية، وفقرات تتعلق باستخدام الحاسوب، شارك في الاختبار تسعة عشر مركز اختبار في خمس ولايات حيث بلغ عدد المشاركين (20,16) مشارك وقد تم استخدام تحليل الانحدار لتحليل البيانات. أظهرت النتائج أنه بالرغم من ضيق مدى تطبيق الدراسة وصغر حجم العينة نسبياً إلا أن هذه الدراسة تقدم أدلة على أن إجراء الاختبارات المحوسبة ممكن عمليًا وبالرغم من أن هناك تحديات تواجه عملية التطبيق إلا أنه يمكن وضع حلول لها لتفاديها، كما لم يكن هناك فروق دالة إحصائياً بين المفحوصين فيما يتعلق باستخدام

الحاسوب ومدى تفضيل جهاز الحاسوب، أو من حيث تصوراتهم حول سهولة استخدامه حسب العمر أو الجنس أو العرق، وبالرغم من أن نسبة من يفضلون اخذ الاختبار بصورته الورقية وصلت إلى 39% إلا أن 75% منهم ممن أخذوا الاختبار بصورة محوسبة ذكروا أنه من السهل أخذ هذه الاختبارات على الحاسوب.

وهدفت دراسة كيم وهيونه (Kim ,Huynh, 2010) إلى البحث في مدى تكافؤ اختبار اللغة الانجليزية المقدم بالقلم والورقة والاختبار الذي يتم تقديمه باستخدام الحاسوب على مستوى ولاية كارولينا للطلبة من ذوي الاحتياجات الخاصة، والطلبة الأصحاء، وقد استندت الدراسة على بيانات من اختبار اللغة الانجليزية للصف التاسع من العام 2008، حيث تم السماح للطلبة بإجراء الاختبار إما باستخدام الحاسوب، أو بالقلم والورقة. بلغ عدد الطلبة الأصحاء 7,000 طالبًا، في حين بلغ عدد الطلبة من ذوي الاحتياجات الخاصة 483 طالباً. وتكون الاختبار من 52 فقرة من نوع الاختيار من متعدد تم تقديم الاختبار المحوسب بنفس ترتيب الاختبار الورقي. وقد أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين الاختبار باستخدام الحاسوب والاختبار بالقلم والورقة.

أما دراسة بياو (Piaw, 2010) فقد هدفت لتطبيق طريقة سالمون لأربع مجموعات تجريبية من أجل: حساب قيمة الثبات لشكلي الاختبار، تحديد ومقارنة تأثير الاختبار عبر شكلي تقديمه، استقصاء أثر شكل الاختبار على: درجات الاختبار، زمن الاختبار، ودافعية الطلاب نحو الاختبار. تكونت عينة الدراسة من (120) أستاذاً ماليزيًا من معهد تدريب المعلمين في وسط شبه جزيرة ماليزيا (41) منهم من الذكور (79) من الإناث بمتوسط عمر (19.1) سنة، تم اختيارهم عشوائيًا من مجتمع المعلمين البالغ (548) معلمًا ومعلمة ممن لديهم نفس مستوى

المهارات على الكمبيوتر من نفس الخلفية التعليمية والتاريخية، قسمت عينت الدراسة إلى أربع مجموعات عشوائية بحجم (30) مشاركًا في كل منها مجموعتان تجريبيتان، ومجموعتان ضابطتان وقد تم ضبط عامل النضج وعامل التاريخ باستخدام اختبارات نفسية بدلا من الاختبارات التحصيلية؛ للتقليل من أثر هذه العوامل لأن الاختبارات النفسية أكثر اتساقا مع مرور الوقت، وتأثير هذه العوامل عليها أقل تم تطبيق اختبار يانبو للأنماط العقلية (YBRAINS) الذي يتكون من (25) فقره يغطيان نفس المحتوى من نوع الاختيار من متعدد بشكلين ورقي والكتروني بحيث تستجيب المجموعتان التجريبيتان على النسخة الالكترونية في حين تستجيب المجموعتان الضابطتان على النسخة الورقية،. بالإضافة إلى تطبيق استبانه مكونة من (54) فقرة تقيس الدافعية على أحد عشر بعدًا تشمل: التحدي، الفاعلية، الفضول، المشاركة، المتعة، تجنب العمل، الاجتماعية، المنافسة، الدرجات، الامتثال، والأهمية . أشارت النتائج أن شكل الاختبار يؤثر بشكل دال إحصائيا على زمن الاختبار والدافعية على النسخة الورقية ويعمل على تقليل زمن الاختبار ويعمل على زيادة الدافعية للمفحوصين وأن النسخة الالكترونية كانت أكثر ثباتًا سواء على مستوى الصدق الداخلي أو الخارجي.

أما دراسة المومني (2012) فقد هدفت إلى بيان أثر الطريقة التي يقدم فيها اختبار اللغة الانجليزية للصف الخامس الأساسي على تقديرات القدرة وخصائص الاختبار وخصائص فقراته السيكومترية. ولتحقيق هدف الدراسة تم تطبيق اختبار في اللغة الانجليزية للصف الخامس الأساسي مكون من (30) فقرة من نوع الاختيار من متعدد، حيث قدم الاختبار في ثلاث طرق مختلفة: الطريقة الأولى(القلم والورقة) قدم فيها الاختبار بالطريقة التقليدية باستخدام القلم والورقة، الطريقة الثانية (حاسوب نموذج1) قدم فيها الاختبار على الحاسوب باستخدام نموذج

واحد، الطريقة الثالثة (حاسوب نموذج 2) قدم فيها الاختبار على الحاسوب باستخدام خمسة نماذج للاختبار تم فيها تغيير ترتيب الفقرات عشوائياً في كل نموذج. تكون مجتمع الدراسة من (2,708) طالباً وطالبة من طلبة الصف الخامس الأساسي في المدارس الحكومية التابعة لمحافظة عجلون في العام الدراسي 2010/2011. تكونت عينة الدراسة من (600) طالباً وطالبة يدرسون في (21) شعبة مختارة بالطريقة العشوائية العنقودية حيث تم توزيع هذه العينة إلى (3) مجموعات تبعاً للطريقة التي قدم فيها الاختبار. تم استخدام البرامج الحاسوبية Bilog, Noharm, LDID في تحليل البيانات الإحصائية. أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية في دالة المعلومات التي يقدمها الاختبار عند مستوى الدلالة (α=0.05)، وعدم وجود فروق ذات دلالة إحصائية في معاملات الثبات الامبريقي بين الطرق الثلاث لتقديم الاختبار عند مستوى الدلالة (α=0.05)، كما أظهرت النتائج وجود علاقة بين مستوى القدرة للطلبة وبين طريقة تقديم الاختبار لصالح الطريقة الثالثة، وعدم وجود تغير في فترات الثقة حول معلمة الصعوبة، وفق نموذج راش، يعزى لاختلاف الطريقة التي يقدم فيها الاختبار.

ثالثًا: الدراسات التي تناولت الأداء التفاضلي حسب شكل تقديم الاختبار (ورقي، محوسب):

أجرى شوارتز وريتش وبودرابسكي (Schwarz, Rich & Podrabsky, 2003) دراسةً هدفت إلى الكشف عن الأداء التفاضلي للفقرة تبعاً لشكل تقديم الاختبار (ورقي، محوسب) من خلال فحص الفروق في الأداء على شكلي تقديم الاختبار، ولتحقيق ذلك تم تطبيق أداتين: الأداة الأولى عبارة عن اختبار قدرات معياري المرجع يدعى (In View) يتكون من عدة اختبارات فرعية في الاستنتاج الكمي، الاستدلال اللفظي، المقارنات، والمتتاليات، يتكون كل

اختبار فرعي منها من (20) فقرة تقدم للطلبة من الصف الرابع إلى الصف التاسع بشكلين ورقي ومحوسب.

طبق الاختبار على عينة تكونت من (12, 195) طالباً وطالبة من الصفوف من الرابع وحتى التاسع) بشكله الورقي ومن(10, 575) طالباً وطالبة من الصفوف من الرابع وحتى التاسع) بشكله المحوسب، وأثناء الاستجابة على الاختبارات المحوسبة كان هناك دراسة مسحية للمفحوصين تتعلق بأُلفة المفحوصين بأجهزة الحاسوب وتصوراتهم عن الاختبارات المحوسبة. الأداة الثانية كانت عبارة عن اختبار في المهارات التعليمية الأساسية للكبار (TABE) وهو اختبار معياري المرجع يقيس المعارف والمفاهيم الأساسية في القراءة، الرياضيات، واللغة عبر فقرات محددة الاستجابة مكون من خمس مستويات (محو الأمية، سهل، متوسط، صعب، متقدم) قدم للمفحوصين بشكلين ورقي ومحوسب، طبق الاختبار على عينة تكونت من 32,12 طالبًا وطالبة بشكله الورقي ومن 426 طالبًا وطالبة بشكله المحوسب، كما أجريت دراسة مسحية للمفحوصين تتعلق بأُلفة المفحوصين بأجهزة الحاسوب وتصوراتهم عن الاختبارات المحوسبة.

تم تحليل بيانات الدراسة باستخدام إحصائيين للكشف عن الأداء التفاضلي طريقة لين هيرنتش(Linn–Harnisch,1981) ، وفرق الأوساط اللامعلمية المعياري (nonparametric standardized mean difference). كشفت النتائج عن وجود فقرتين أظهرتا أداءً تفاضلياً على اختبار(TABE)، حيث كانت الفروق ما بين المجموعات على الاختبار المحوسب والورقي أكثر وضوحًا عند ذوي القدرات المنخفضة، كما أن أراء الطلاب كانت محايدة حول درجة تفضيلهم للاختبارات المحوسبة، بالإضافة إلى أن طريقة الأداء التفاضلي، تمثل طريقة مهمة لدراسة الفروق بين المجموعات على مستوى الفقرة.

وفي دراسةٍ أجراها جو ودراك وولف (Gu, Drake & Wolfe, 2006) هدفت إلى معرفة فيما إذا كانت صعوبة الفقرات تختلف باختلاف طريقة تقديم الاختبار ورقي ومحوسب عن طريق إجراء تحليل للأداء التفاضلي متبوع بعمل تحليل مفصل لمحتوى الفقرات، من خلال تطبيق ثلاثة اختبارات متكافئة في الرياضيات بشكلين الكتروني وورقي، الشكل الورقي مكون من (20) فقرة اختيار من متعدد على كل نموذج، أول 12 فقرة خاصة بالمقارنات الكميه وما تبقى من الفقرات تتعلق بحل المشكلات وقد تم معادلة شكلي الاختبار بمعامل ثبات 0.65 على الشكل الورقي و 0.62 على الشكل الالكتروني، وقد تكونت عينة الدراسة من (165) طالباً: (55%) منهم من الذكور و(45%) من الإناث، تم اختيارهم من تشكيلة واسعة من التخصصات الأكاديمية (30% تقريبًا من العلوم الطبيعية، 30% من الهندسة، 15% من العلوم الاجتماعية والباقي من مجموعه متنوعة من التخصصات).

كشفت النتائج أن 38% من الفقرات أظهرت أداء تفاضليًا لصالح الشكل الالكتروني لذلك فقد تم عمل تحليل محتوى تم التركيز فيه على خصائص الفقرات من حيث تنسيق الصفحة، ملاحظات رياضية (إشارات العمليات الحسابية، متغيرات، أشارت المقارنات، مسائل وتمارين رياضية مكونه من جمل نصية) المحتوى الرياضي للفقرات (حساب، هندسة، حل معادلات، واختيار أرقام) بالرغم من أن النتائج أظهرت أن الاختلاف في تنسيق الصفحة، وكيفية الاستجابة على الفقرة على الشكل الالكتروني كانت مسئوله بشكل قليل عن الأداء التفاضلي على فقرات هذا الشكل إلا أن الاختلافات في الملاحظات الرياضية الواردة في نص الفقرة والاختلاف في المحتوى الرياضي للفقرات كان لها علاقة قويه مع الأداء التفاضلي على فقرات الشكل الالكتروني.

وأجرى بوهن وبوتن وكيم (Puhan, Boughton & Kim, 2007) دراسةً هدفت إلى مقارنة أداء الطلبة على امتحان القبول في البرامج التعليمية المتكون من نسختين إحداها ورقية والأخرى الكترونية لقياس المهارات الأساسية في مجالات القراءة، والكتابة، والرياضيات حيث تتكون كلا النسختين من (40) فقرة من نوع الاختيار من متعدد في مجال القراءة و(40) فقرة من نوع الاختيار من متعدد في مجال لرياضيات و(45) فقرة بعضها اختيار من متعدد وبعضها مقالية في مجال الكتابة، تم تطبيق نسختي الاختبار على ست مجموعات من المفحوصين تم تقسيمهم بالاعتماد على شكل الاختبار (الكتروني، ورقي) وبالاعتماد على مجالات الاختبار (قراءة، كتابة، رياضيات) وأعطي الطلاب الحرية للاختيار بين النسخة الالكترونية والنسخة الورقية لضمان تساوي قدرات المفحوصين وكانت أحجام العينات كالتالي: على النسخة الالكترونية على مستوى الاختبار ككل (11, 22 في مجال القراءة ,10, 50 في مجال الكتابة ,11, 36 في مجال لرياضيات)، وعلى مستوى فقرات الاختبار (11,22 في مجال القراءة، 10, 50 في مجال الكتابة ، 11, 36 في مجال لرياضيات). أما على النسخة الورقية فد تكونت العينة من مجموعتين فرعيتين عشوائيتين تتكون كل منهما من (2000) مفحوص على مجالات كل من القراءة، الكتابة والرياضيات على التوالي، وقد تم تحليل البيانات باستخدام مقياس حجم الأثر المعروف بمقياس كوهنز d (Cohen, 1988) وتم عمل تحليل للأداء التفاضلي للفقرات (DIF) باستخدام طريقة (SIBTEST) لمقارنة الأداء على مستوى الاختبار ككل.

دلت النتائج على أن قيمة حجم الأثر غير دالة إحصائيا حيث كانت قليلة(d < 0.20) مما يشير إلى عدم وجود فروق كبيره في الأداء على نسختي الاختبار، كما كشف تحليل الأداء التفاضلي للفقرات على أن الأداء على فقرات القراءة والكتابة قابلة للمقارنة على كلا النسختين،

وأن هناك ثلاثة فقرات أظهرت أداء تفاضليا لصالح الشكل المحوسب، فشلت المراجعة الموضوعية من قبل المحكمين في تحديد فيما إذا كانت الفروق في الأداء يمكن تفسيرها نتيجة الاختلاف في شكل الاختبار (ورقي أو الكتروني) لذلك لم يُستطع تحديد الأسباب المسئولة عن الأداء التفاضلي.

تعقيب على الدراسات السابقة:

في ضوء استعراض الدراسات والأدب التربوي، وبعد البحث في الدوريات والمواقع الالكترونية، الرسائل الجامعية، الكتب التربوية لوحظ ما يلي:

الدراسات التي تناولت اثر طريقة تقديم الاختبار على: الدافعية، مدى تفضيل الاختبارات المحوسبة، الاتجاهات نحوها، ألفة الطلبة بها، وعلاقة ذلك بالأداء على فقرات الاختبار، أعطت نتائجها هذه الدراسة ثقة بأن يتم البحث دون قلق من أن تكون نتائج هذه الدراسة ناتجة عن عدم تقبل الطلبة للاختبارات المحوسبة، وليس عن طريقة تقديم الاختبار، حيث أن نتائج هذه الدراسات تشير إلى أن مدى اختلاف أداء المشاركين على شكلي تقديم الاختبار، يعتمد على مدى ألفة الفرد باستخدام الحاسوب، وأن مدى ألفة الفرد بالحاسوب تستطيع أن تتنبأ بشكل دال إحصائياً بأداء المفحوصين على الاختبارات المحوسبة، كما في دراسة هوركاي وبينت والين وكابلان ويان (2006) ، وأن شكل تقديم الاختبار يؤثر بشكل دال إحصائيًا على زمن الاختبار والدافعية، ويعمل على تقليل زمن الاختبار، وزيادة الدافعية للمفحوصين، كما في دراسة بياو (2010).

معظم الدراسات العربية والإقليمية والعالمية التي تناولت موضوع الأداء التفاضلي على مستوى اختبارات وطنية عامة، كانت تدرس أثر متغيرات مثل: الجنس، والعرق، وحجم العينة، واللغة، وعدد فقرات الاختبار...الخ، وليس أثر شكل تقديم الاختبار على الأداء التفاضلي؛ لذلك فإن هذه الدراسة جاءت لتضيف متغيرًا جديدًا، لتحري الأداء التفاضلي يتعلق بشكل تقديم الاختبار (ورقي، محوسب)، لم تضفه الدراسات السابقة، قد يكون له أثر في تطوير فقرات اختبارات تتصف بالعدالة والمصداقية في قياس الأداء الحقيقي للطالب.

هنالك ندرة في الدراسات التي تهدف إلى معرفة أثر شكل تقديم الاختبار(ورقي، محوسب) على خصائص الاختبار وفقراته سواء على المستوى العربي، أو على المستوى المحلي. كما أن الدراسات في هذا الصدد على ندرتها، لم تتناول الأداء التفاضلي كأحد النتائج الناجمة عن هذا الأثر، حيث لا يوجد أي دراسات عربية أو محلية تناولت هذا الجانب كهدف رئيسي لدراسة شاملة على مستوى اختبار وطني، لذلك جاءت هذه الدراسة، لتحري الأداء التفاضلي بشكل شمولي لاختبار وطني تبعًا لطريقة تقديم الاختبار، في محاولة للحصول على معلومات يتم من خلالها المشاركة في تقديم المشورة والتوصيات، لجعل عملية التوسع في تطبيق الاختبارات المحوسبة، تراعي العدالة والمساواة بين الطلبة.

كما أن معظم الدراسات التي تناولت شكل تقديم الاختبار سواء على المستوى المحلي أو الإقليمي أو العالمي، كانت تركز على دراسة هذا الأثر من حيث: الفروق في الأداء على درجات الاختبار، ومدى تكافؤ الاختبارات، وأثر شكل تقديم الاختبار على خصائص الاختبار وفقراته، وبالرغم من أن هذه الدراسات لم تبحث في تحري الأداء التفاضلي لفقرات الاختبار تبعًا لطريقة تقديمه بطريقة مباشرة كهدف رئيسي، إلا أنها تشير ضمنياً إلى أن الأداء التفاضلي قد يكون من ضمن الأسباب الكامنة وراء الدلالة الإحصائية للفروق بين شكلي الاختبار على

المتغيرات السابقة، من خلال التفاعل ما بين شكل تقديم الاختبار وعضوية المجموعة، لذلك أوصت هذه الدراسات بمزيد من الأبحاث، من أجل دعم مدى صدق وموثوقية الاختبارات المحوسبة، من خلال تسليط الضوء على دراسة العوامل التي تتفاعل مع قدرة المفحوصين على الإجابة عن فقرات الاختبار، إضافةً إلى محتوى الفقرة مثل: عوامل هيكلة شكلي الاختبار وصور الإخراج لهما، وخصائص المفحوصين، التي تسهم بالعديد من الآثار الناتجة عن طريقة تقديم الاختبار، وخاصة عدم جدية المفحوصين، مما يؤدي إلى فروق في احتمال الإجابة الصحيحة على الفقرة أو الاختبار عبر المجموعات المتساوية بالقدرة تعود لطريقة تقديم الاختبار. واستجابة لهذه التوصيات جاءت هذه الدراسة، لتتناول هذه العوامل بشيء من التركيز.

الدراسات التي أجريت على المستوى العالمي في ما يتعلق بتحري الأداء التفاضلي للفقرة تبعًا لشكل تقديم الاختبار كانت محدودة أيضًا، منها دراسة شوارتز وريتش وبورابسكي (2003)، التي أشارت نتائجها إلى أن الفقرات التي أظهرت أداءً تفاضليًا كانت لصالح الشكل المحوسب، وأن الفروق كانت أكثر وضوحًا عند مستويات القدرة المنخفضة، إما دراسة جو ودراك وولف (2006)، فقد أشارت نتائجها إلى أن الأداء التفاضلي كان لصالح الاختبار الالكتروني، وأن الاختلاف في تنسيق الصفحة، وكيفية الإجابة عن الفقرة مسئولة بشكل قليل عن ظهور الأداء التفاضلي، إلا أن الاختلافات في الملاحظات الرياضية الواردة في نص الفقرة، والاختلاف في المحتوى الرياضي للفقرات كان لها علاقة قوية مع الأداء التفاضلي على فقرات الشكل الالكتروني.

في حين أشارت نتائج دراسة بوهن وبوتن وكيم (2007)، إلى أن الأداء على فقرات القراءة والكتابة قابلة للمقارنة على نسختي الاختبار، بالرغم من وجود الأداء التفاضلي، ولكنها لم تستطع أن تثبت فيما إذا كانت الفروق في الأداء ناتجة عن اختلاف شكل تقديم الاختبار، ولم

تستطع تحديد الأسباب المسئولة عن الأداء التفاضلي؛ لذلك فإن هذه الدراسة جاءت لتكمل ما بدأه الآخرون لإثراء الأدب التربوي كميًا ونوعيًا في هذا المجال، وللتأكد من أن نتائج استخدام التكنولوجيا، والاستفادة من مزاياها، في التوسع في تطبيق الاختبارات المحوسبة يمكن الدفاع عنه بقوة، وأن مزايا هذه الاختبارات تفوق أوجه القصور فيها، وأنه إذا كان هناك فروق في الأداء تعود إلى وجود الأداء التفاضلي على فقرات الاختبار، فكيف يمكن بناء اختبارات محوسبة، بشكل يحد من هذه الاختلافات على طريقتي تقديم الاختبار خاصة وأن الدراسات التي تبحث في تحري الأداء التفاضلي، تهدف إلى مساعدة مطوري الاختبارات ومن يقوم بكتابة الفقرات على التخلص، أو على الأقل الحد من هذه الفقرات.

كما تستخدم هذه الدراسة طريقة DFIT للكشف عن الأداء التفاضلي على مستوى الفقرة والاختبار ككل من خلال مؤشرات خاصة بذلك وقد أشير إلى هذه الطريقة بشكل مفصل في الإطار النظري للدراسة، مما قد يكون له أثر في تحقيق هدف الدراسة، في إجراء تحري شامل ودقيق، لمعرفة فيما إذا كانت الاختبارات المحوسبة بديل صادق وعادل عن الاختبارات الورقية.

الفصل الثالث

الطريقة والإجراءات

في هذا الجزء يكون الحديث عن اداة الدراسة، وبيانات الدراسة التي تم الحصول عليها، والطريقة والإجراءات، والأساليب الإحصائية التي تم استخدامها للإجابة على أسئلة الدراسة.

بيانات الدراسة

يتكون مجتمع الدراسة من جميع طلاب وطالبات الصف العاشر الأساسي للعام الدراسي 2011/2010، ممن تقدموا إلى اختباري العلوم والرياضيات على النسختين الورقية والمحوسبة في مدارس المملكة الأردنية الهاشمية الحكومية والخاصة، والمدارس التابعة لوكالة الغوث الدولية، والمدارس التابعة لمديرية التعليم والثقافة العسكرية، حيث يتم تطبيق الاختبار على جميع الطلبة والطالبات من قبل وزارة التربية والتعليم، ومبرر هذا الاختيار، أن الصف العاشر يمثلُ نقطة انعطاف، لكونه أخر صفوف المرحلة الأساسية، الذي يُتخذ بناءً على نتائجه قرارات هامة، بتصنيف الطلبة وتوزيعهم على فروع مرحلة التعليم الثانوي، ولكون المنهاج الأردني منهاج حلزوني تكاملي، تبنى فيه المعرفة بشكل تراكمي، لذلك فالنتاجات التعليمية المتحققة في نهاية هذا الصف، عبارة عن مجموع النتاجات التعليمية المتحققة في الصفوف التعليمية السابقة له، وبما أن الهدف من الاختبار الوطني، قياس مدى امتلاك الطلبة لمهارات التعلم الأساسية، فإن الحلول التي توضع لمعالجة نقاط الضعف، عبارة عن حلول شاملة لمجموع نقاط الضعف عبر المرحلة الأساسية برمتها، بالإضافة إلى أن الطلبة يكونون أكثر جدية في التعامل مع الاختبار، وأكثر قدرة على محاكمة الأمور، لفهم وتنفيذ تعليمات الاختبار، مما يجعل النتائج أكثر واقعية ودقة في قياس السمة المنشودة من الاختبار.

وبعد الرجوع إلى ملف البيانات الذي تم الحصول عليه من قسم الاختبارات التابع لوزارة التربية والتعليم، تم اختيار عينة للدراسة مقدارها 2800 طالباً وطالبةً بشكل عشوائي من مجتمع الدراسة، ممن تقدموا على اختباري العلوم والرياضيات بشكليه الورقي والمحوسب، بواقع 1400 طالباً وطالبة عبر اختبار العلوم، و1400 طالباً وطالبة عبر اختبار الرياضيات، ومبرر هذا الاختيار، الاستجابة للمحددات الخاصة بكل من النموذج الثلاثي المعلمة ضمن نظرية الاستجابة للفقرة، والطريقة المستخدمة في الدراسة، لغاية الكشف عن الأداء التفاضلي للاختبار وفقراته، لضمان أن تأتي النتائج في خدمة أغراض الدراسة، ويبين الجدول (1) توزيع عينة الدراسة.

جدول (1)

توزيع عينة الدراسة على اختباري العلوم والرياضيات من الاختبار الوطني لضبط نوعية التعليم بشكليه الورقي والمحوسب.

طريقة تقديم الاختبار	الاختبار	عدد الطلبة	المجموع
ورقي	علوم	700	1400
	رياضيات	700	
محوسب	علوم	700	1400
	رياضيات	700	

أداة الدراسة:

تكونت أداة الدراسة من "الاختبار الوطني لضبط نوعية التعليم" للصف العاشر الأساسي للعام الدراسي 2011/2010 المقدم من قبل وزارة التربية والتعليم/إدارة الامتحانات والاختبارات المشار له في الإطار النظري من هذه الدراسة، وهو اختبار تحصيلي من نوع (الاختيار من متعدد)، لكل فقرة منها أربعة بدائل، واحدة منها صحيحة تأخذ فقرات الاختبار جميعها إحدى العلامتين (0 ,1)، وتم اختيار اختباري العلوم والرياضيات من ضمن رزمتين اختباريتين:

الأولى في (اللغة العربية والرياضيات)، والثانية في (اللغة الإنجليزية والعلوم) كأداة للدراسة؛ لغايات الكشف عن الأداء التفاضلي، ومبرر هذا الاختيار رفض مديرية الاختبارات إعطاء البيانات الخاصة بكافة مجالات الاختبار، لأسباب إدارية وأسباب تتعلق بسرية البيانات. تكونت أداة الدراسة لاختباري العلوم والرياضيات من:

- الشكل الورقي: يتم تطبيقه بصورة يدوية، باستخدام الورقة وقلم الرصاص.
- الشكل المحوسب: ويتم تطبيقه باستخدام الشبكة الإلكترونية المحلية للمدارس (Etesting.Elearning.jo)، في نفس الوقت الذي يتم فيه تطبيق الشكل الورقي للاختبار.

تكون اختبار العلوم بشكليه من (30) فقرة من نوع الاختيار من متعدد، لكل فقرة منها أربعة بدائل موزعة على محاور: الأحياء، الفيزياء، علوم الأرض والبيئة، الكيمياء. تقيس هذه الفقرات أداء الطلبة على مهارات تتعلق بـــ (توظيف الاستقصاء في التوصل إلى وراثة صفة عند الإنسان، العوامل المؤثرة في نمو البكتيريا، دور النبات في الحفاظ على التوازن البيئي، التوصل للخصائص الفيزيائية والكيميائية لبعض الفلزات وظاهرة شذوذ الماء. توظيف العلاقات الرياضية في حل المسائل حول (المرايا، العدسات، مقاومة الموصل، قانون سنِل). قراءة الأشكال والجداول والرسومات البيانية المتعلقة بـــ (خرائط الطقس، الجدول الدوري...)، كتابة صيغ أولية وجزيئية وبنائية، ومعادلات موزونة، ورموز لويس لبعض الذرات، وبنى لويس لبعض الجزيئات وكتابة طرز جينية...)، ويبين الجدول (2) توزيع أسئلة اختبار العلوم حسب المحاور التي يتكون منها. وللاطلاع على مضمون فقرات الاختبار انظر الملحق (1)

الجدول (2)

توزيع أسئلة العلوم بحسب المحور

المحور	عدد الفقرات	النسبة المئوية%	أرقام الفقرات حسب المحاور
أحياء	8	26.66	8,7,6,5,4,3,2,1
فيزياء	9	30	17,16,15,14,13,12,11,10,9
علوم ارض والبيئة	5	16.66	30,29,28,27,26
كيمياء	8	26.66	25,24,23,22,21,20,19,18
المجموع	30	100	

كما تكون اختبار الرياضيات على شكلي الاختبار من (30) فقرة من نوع الاختيار من متعدد، لكل فقرة منها أربعة بدائل، موزعة على محاور: الجبر، الهندسة والقياس، والإحصاء والاحتمالات. تقيس هذه الفقرات أداء الطلبة على المهارات الآتية: كتابة قاعدة الاقتران الحقيقي وتمثيله بيانيًا، إيجاد الاقتران العكسي، وناتج تركيب اقترانيين. حل أنظمة المعادلات، إجراء العمليات على المصفوفات، تطبيق الخصائص الهندسية للدوائر والمضلعات، والقوانين الأساسية في الهندسة التحليلية. إيجاد النسب المثلثية للزوايا الدائرية، إيجاد المقاييس الإحصائية وأثر التحويلات الخطية فيه، إيجاد احتمالات الحوادث البسيطة والمشروطة، ويبين الجدول (3) توزيع أسئلة اختبار الرياضيات. وللاطلاع على مضمون فقرات الاختبار انظر الملحق (2)

الجدول (3)

توزيع أسئلة الرياضيات بحسب المحور

المحور	عدد الفقرات	النسبة المئوية%	أرقام الفقرات حسب المحاور
جبر	12	40	17,16,15,14 ,13 , 12 , 11 , 6 , 4 , 3 , 2 , 1 ,
هندسة وقياس	10	33.33	22,21,20,19,18,10,9,8,7,5
إحصاء واحتمالات	8	26.26	30,29,28,27,26,25,24,23
المجموع	30	100	

خطوات إعداد الاختبار من قبل إدارة الاختبارات في وزارة التربية والتعليم:

شُكِّلت أربع لجان، لكتابة فقرات اختبارات الصف العاشر في الرياضيات والعلوم واللغة العربية واللغة الإنجليزية، وتضمنت هذه اللجان مشرفين تربويين من إدارة الامتحانات والاختبارات، وأعضاء من إدارة المناهج والكتب المدرسية، ومشرفين تربويين من إدارة التدريب والتأهيل والإشراف التربوي، وعددًا من مشرفي الميدان، وقد أُتبعت في بناء الاختبارات الخطوات الآتية:

- **بناء الفقرات**

قامت اللجان بدراسة نتاجات التعّلم للصف العاشر في مباحث الرياضيات والعلوم واللغة العربية واللغة الإنجليزية، وتحديد مهارات التعّلم الأساسية التي سيقيسها كل اختبار من الاختبارات، ثم عملت على إعداد جدول مواصفات لكل اختبار، وبناء تجمعات من الفقرات الإختبارية، وطباعتها في رزمتين اختباريتين: الأولى في (اللغة العربية والرياضيات)، والثانية في (اللغة الإنجليزية والعلوم).

- **تجريب الفقرات**

جُربت الفقرات على عينة من الطلبة، وأجري التحليل الإحصائي لإجابات الطلبة عن الفقرات، للتعرف على خصائصها السيكومترية مثل: فاعلية البدائل، مستوى صعوبة الفقرات، تمييزها، ومقرؤيتها، ثم اختيرت الفقرات الملائمة، وعدلت صياغة بعض الفقرات، وحذفت الفقرات التي كان تمييزها سالبًا أو قريبًا من الصفر.

- **إعداد الورقة الاختبارية بصيغتها النهائية**

انتُقيت فقرات الاختبار من تجمع الفقرات، بحيث كان انتقاء الفقرات مُمثلا للأوزان والنسب في جدول المواصفات.

- **إيجاد الخصائص السيكومترية للاختبارات:**

أ- الصدق: حُلل محتوى الاختبارات، للتأكد من صدق المحتوى (Content Validity) وذلك بتحديد النتاج الذي تقيسه كل فقرة، والوحدة أو الفصل الذي ينتمي له ذلك النتاج، ووزن العلامة المخصصة لتلك الفقرة، ومن ثم تم مطابقة جدول تحليل الاختبار بجدول المواصفات.

ب- ثبات الاختبار: حُسب معامل ثبات كل اختبار باستخدام معادلة "كرونباخ ألفا"، حيث كانت على النحو الآتي (وزارة التربية والتعليم، 2011):

جدول (4)

معامل الثبات"كرونباخ ألفا" لمجالي العلوم والرياضيات المستمدين من الاختبار الوطني عبر شكلي الاختبار (ورقي، محوسب)

طريقة تقديم الاختبار	المبحث	معامل الثبات
ورقي	علوم	.682
	رياضيات	.819
محوسب	علوم	.711
	رياضيات	.826

إجراءات تطبيق الاختبار:

- عُقد اجتماع مع رؤساء أقسام الامتحانات والاختبارات في مديريات التربية والتعليم كافة، وتم توضيح أهداف الاختبار الوطني لضبط نوعية التعليم، وإجراءات تطبيقه، وموعد تسّلم مغلفات الأسئلة وأوراق الإجابة وتسليمها.

- قام مدراء التربية والتعليم بتكليف رؤساء أقسام الامتحانات والاختبارات، بإعداد كشف يتضمن أسماء رؤساء القاعات والمراقبين المكلفين بتطبيق الاختبار، حسب أعداد المدارس شريطة أن يكون رئيس قاعة واحد لكل مدرسة، ومراقب واحد لكل شعبة يطبق فيها الاختبار الورقي، على أن يتم تكليف قيم المختبر/معلم الحاسوب كمراقب على تطبيق الاختبار المحوسب.

- سُلمت مغلفات الأسئلة وأوراق الإجابة لرؤساء أقسام الامتحانات والاختبارات من إدارة الاختبارات والامتحانات حسب الكتاب الرسمي، بحيث كانت تحتوي على:

  - نماذج الإجابة (القارئ الضوئي).
  - تقرير سير الاختبار من كل مدرسة.
  - نموذج رقم (1)، المخصص لأسماء الطلبة الأصليين معبأ حسب الأصول.
  - نموذج رقم (2)، المخصص لأسماء الطلبة الإضافيين معبأ حسب الأصول.

- قام كل من: مدير الشؤون التعليمية والفنية، رئيس قسم الامتحانات والاختبارات، رئيس قسم الشبكات وصيانة الحاسوب كل في مديريته بعقد اجتماع لرؤساء القاعات، وضحت به آلية تطبيق الاختبار الورقي والمحوسب بطريقة سليمة بعد أدائهم القسم بصوت مسموع، ثم تسليمهم مغلفات الأسئلة مع نماذج الإجابة (القارئ الضوئي) للاختبار الورقي حسب عدد طلبة المدارس مغلقة، مع التأكيد على ضرورة عدم فتحها إلا وقت الاختبار بوجود المراقبين المعنيين بالتطبيق، وذلك تحت طائلة المسائلة القانونية، وضرورة تسليم نماذج

الإجابة مرتبة حسب الأوراق مع الكشوف والتقارير الخاصة بتطبيق الاختبار معبأة حسب الأصول، بعد تطبيق الاختبار مباشرة.

- وقبل تطبيق الاختبار قامت مديريات التربية في مختلف مناطق المملكة بعقد ورش تدريبية لمدراء المدارس، قيمي مختبرات الحاسوب/معلمي الحاسوب بالمدارس التي لا تحتوي على قيمي مختبرات، والمعلمين المعنيين بتطبيق الاختبار، من قبل مشرفين تربويين، تم خلالها تدريبهم على كيفية تطبيق الاختبار، تعليمات الاختبار وأهميته كواجب وطني.

- قام قيمي مختبرات الحاسوب/معلمي الحاسوب بعد أداء القسم بتسلم مغلفات تحتوي على بيانات الدخول لحساب مدارسهم على موقع الاختبارات الإلكترونية، حيث تم التأكد من صحة البيانات الأساسية للمدارس، وتعديل تلك البيانات إذا لزم، ثم تم إنشاء حسابات للطلبة، والاطلاع على الملف التدريبي الموجود على الموقع الذي يوضح آلية العمل خطوة بخطوة، والتأكد من جاهزية مختبرات الحاسوب، لعقد الاختبار المحوسب وخصوصًا عدد الحواسيب الصالحة المتصلة بالشبكة، حيث تم الاستعانة بفريق صيانة شبكات الحاسوب كلٌ في مديريته حسب الحاجة لذلك.

- قام كل قيم مختبر/معلم الحاسوب بالتنسيق مع مدير مدرسته باختيار الطلبة الذين سيتقدمون للاختبار المحوسب من الصف العاشر، بحيث كان عددهم مطابق لعدد حسابات الطلبة المتاحة على حساب المدرسة، وتم اختيار الشعب بشكل عشوائي من الصف العاشر، بما يتلاءم وعدد الحواسيب بالمدرسة.

- قام كل قيم مختبر/معلم الحاسوب بالدخول إلى حساب مدرسته، وتم إدخال أسماء الطلبة المعنيين بالاختبار المحوسب، وأرقامهم الوطنية، وشعبهم في المكان المخصص بجانب رمز المستخدم وكلمة المرور (إدخال أصفار للطلاب الغير أردنيين)، ثم قاموا بتصدير حسابات الطلبة إلى ملف ورد (word) أو اكسل (excel)، وطباعتها ورقيًا، وتوزيعها على الطلبة قبل بدء الاختبار، للدخول إلى الاختبار.

- قام قيمي المختبرات/معلمي الحاسوب في المدارس التي تُطبِق الاختبار بشكله المحوسب بتدريب الطلبة على كيفية الدخول إلى الموقع الإلكتروني المعد لذلك، وكيفية الإجابة عن فقرات الاختبار، من خلال فقرات تجريبية.

- قام كل رئيس قاعة بعقد اجتماع للمراقبين قبل موعد الاختبار بساعة واحدة على الأقل، حيث قام المراقبين بأداء القسم أمامه، وبين رؤساء القاعات للمراقبين أهمية الاختبار كواجب وطني، ثم قام المراقبين مع رؤساء القاعات بتفقد قاعات الاختبار من حيث، النظافة، التهوية، الإضاءة ومناسبة المقاعد مع أعداد الطلبة، كما تم تصوير الصفحات الناقصة في المدارس التي صادف وجود خلل ما، أو نقص في إحدى كراسات الاختبار، أو العدد المناسب، وعدا ذلك كان يدون في بند (الصعوبات الفنية والإدارية) من تقرير سير الاختبار (وزارة التربية والتعليم، 2011).

- طبق الاختبار بشكليه الورقي والمحوسب يوم الاثنين الموافق 2011/5/16 الساعة الحادية عشرة صباحاً في جميع مدارس المملكة التي تضم الصف العاشر، بتقديم رزمتين اختباريتين: الأولى في (اللغة العربية والرياضيات)، والثانية في (اللغة الإنجليزية والعلوم)،

وكانت مدة الاختبار ساعة ونصف. (مدير إدارة الاختبارات في وزارة التربية، اتصال شخصي، 10 تشرين أول، 2012).

- على الشكل الورقي قام مراقبي القاعات بتوضيح أهداف الاختبار، وأهميته للطلبة، وتم إعطاء وقت كاف للطلبة قبل بدء الاختبار، قام الطلبة خلاله بتعبئة المعلومات وتظليل البيانات التالية على نموذج الإجابة (القارئ الضوئي): اسم الطالب، اسم المدرسة، رقم جلوس الطالب (حسب كشوف الطلبة)، جنس المدرسة (تظليل الرقم (1) إذا كان الطالب ذكراً، تظليل الرقم (2) إذا كان الطالب أنثى، تظليل رمز المبحث (تظليل الرقم (83) إذا كان المبحث رياضيات، تظليل الرقم (84) إذا كان المبحث علوم، تظليل الرقم الوطني للمدرسة حسب الأرقام الوطنية الموجودة على كشوف الطلبة، تظليل رمز المديرية حسب الكشف المرفق بالملحق، وتظليل رمز الصف (يظلل الرقم (10))، وتأكد المراقبون من صحة تلك البيانات.

- تم إعطاء تعليمات للطلبة بالمحافظة على نموذج الإجابة (القارئ الضوئي) بعدم ثنيه أو تمزيقه أو الكتابة العشوائية بداخله، والمحافظة على نظافته، وبُينت آلية الإجابة على نموذج الإجابة (القارئ الضوئي)، بحيث يتم تظليل رمز الإجابة الصحيحة باستخدام أقلام رصاص، وعدم استخدام أقلام الحبر، وفي حال تغيير الإجابة على النموذج يجب استخدام الممحاة لإزالة أثر الإجابة السابقة، ثم تظليل الإجابة الجديدة، وأُعطي الطلبة تعليمات بالالتزام بزمن الامتحان المحدد على الورقة الاختبارية، والإجابة عن جميع الأسئلة بعناية ودقة، وأنه بإمكانهم سؤال مراقب القاعة، إذا شعر الطالب بحاجته للمساعدة، لفهم سؤال ما، وتم توزيع الاختبار على الطلبة بشكل متناوب، بحيث يأخذ الطالب الأول الرزمة

الأولى (اللغة العربية ورياضيات)، ويُعطى الطالب الذي يليه الرزمة الثانية (لغة إنجليزية وعلوم) وهكذا.

- في نفس الوقت على الشكل المحوسب قام قيمي المختبرات/معلمي الحاسوب بإجلاس الطلبة على الحواسيب المخصصة للاختبار المرتبة على شكل حرف (U)؛ كإجراء لضبط الغش في الاختبار، ثم وزعت حسابات الدخول (رمز المستخدم وكلمة المرور) على الطلبة، حيث قام كل طالب بإدخال الحساب الخاص به، للدخول إلى موقع الاختبارات الإلكترونية، وبعد دخول الطلبة إلى موقع الاختبارات الإلكتروني (Etesting. Elearning.jo)، قام كل طالب بإدخال البيانات الشخصية الخاصة به كما هي على نموذج القارئ الضوئي الخاص بالشكل الورقي للاختبار؛ إذا لم يقم قيم المختبر/معلم الحاسوب بذلك سابقاً، وتأكد المراقبون من صحة تلك البيانات .
- بعد دخول الطالب إلى موقع الاختبار تظهر له إحدى رزمتي الاختبار إما (لغة عربية ورياضيات، أو لغة إنجليزية وعلوم)، وتظهر له تعليمات الاختبار كما هي عليه بالشكل الورقي؛ إلا أن الإجابة هنا تكون بالضغط على رمز الإجابة الصحيحة باستخدام الفأرة بدلاً من التظليل باستخدام قلم الرصاص، وأن الطالب يعتبر غير متقدم للاختبار إذا لم يقم بالنقر على زر (إنهاء)، ويبدأ الاختبار عند ضغط الطالب على زر (ابدأ)، وقد أشرف قيمي المختبرات/معلمي الحاسوب على الاختبار، بالمراقبة على الاختبار، وتولي الأمور الفنية أثناء تطبيق الاختبار، وزود الطلبة بورق ابيض لاستخدامه عند الحاجة أثناء تطبيق الاختبار (وزارة التربية والتعليم، 2011).
- بعد استلام أوراق القارئ الضوئي، ونتائج الطلبة على الاختبار المحوسب تم إدخال استجابات الطلبة إلى الحاسوب باستخدام القارئ الضوئي، وتصحيح الاختبار آليًا.

- تم معالجة البيانات الخام، وذلك بحذف استجابات الطلبة الذين كانت نسبة الفقرات التي لم يجيبوا عنها (المحذوفة) تزيد على (80%)، وحذف استجابات الطلبة الذين اختاروا البديل نفسه على (80%) أو أكثر من فقرات الاختبار.
- تم إجراء تحليل إحصائي لنتائج الاختبار بصورتيه الورقية والمحوسبة، وإيجاد كل مما يأتي: الوسط الحسابي والانحراف المعياري لأداء الطلبة على مستوى المملكة والمديرية والمدرسة. الوسط الحسابي لأداء الطلبة على محاور التعلّم ومهاراته الأساسية على مستوى المملكة والمديرية والمدرسة. نسبة أداء الطلبة على مستوى المملكة، وعلى مستوى كل مديرية في مستويات الأداء المعتمدة. حُددت مستويات أداء الطلبة على محاور الاختبار ومهارات التعّلم الأساسية في مباحث الرياضيات والعلوم وهي: المستوى المتقدم (85 فما فوق)، والمستوى الماهر (60-84)، المستوى الماهر جزئيًا (35-59)، المستوى المبتدئ (34 فما دون) (وزارة التربية والتعليم، 2011).

الجدول (5)

مستويات أداء الطلبة على محاور اختبار الرياضيات ومهارات التعلم الأساسية فيه.

درجات قطع	المهارات	المستوى	ورقي	محوسب
84	يحقق النتاجات التعليمية بشكل يفوق معايير المستوى التعليمي المحدد.	**مستوى المتقدم** Advanced Exceeding Level	% 4	% 4
59	يحقق النتاجات التعليمية بشكل يلائم معايير المستوى التعليمي المحدد	**مستوى الماهر** Fully Proficient Meeting Level	%13	% 17
34	يحقق الحد الأدنى المطلوب من النتاجات التعليمية لمعايير المستوى التعليمي المحدد	**مستوى الماهر جزئيا** Nearly Proficient Approaching Level	% 37	% 39
	لم يحقق الحد الأدنى المطلوب من النتاجات التعليمية بشكل يلائم معايير المستوى التعليمي المحدد، ويحتاج إلى خطه علاجية لإعادة توجيه تعلمه إلى المسار الصحيح	**مستوى المبتدئ** Basic (Staring) Level	% 46	% 40

الجدول (6)

مستويات أداء الطلبة على محاور اختبار العلوم ومهارات التعلم الأساسية فيه

درجات قطع	المهارات	المستوى	ورقي	محوسب
	يحقق النتاجات التعليمية بشكل يفوق معايير المستوى التعليمي المحدد.	**مستوى المتقدم** Advanced Exceeding Level	%1	%0
84	يحقق النتاجات التعليمية بشكل يلائم معايير المستوى التعليمي المحدد	**مستوى الماهر** Fully Proficient Meeting Level	%12	%16
59	يحقق الحد الأدنى المطلوب من النتاجات التعليمية لمعايير المستوى التعليمي المحدد	**مستوى الماهر جزئيا** Nearly Proficient Approaching Level	%46	%49
34	لم يحقق الحد الأدنى المطلوب من النتاجات التعليمية بشكل يلائم معايير المستوى التعليمي المحدد، ويحتاج إلى خطه علاجية لإعادة توجيه تعلمه إلى المسار الصحيح	**مستوى المبتدئ** Basic (Staring) Level	%41	%35

إجراءات الدراسة:

في تاريخ 2013/2/26 تم الحصول على الموافقات الخطية الرسمية، بناءً على كتاب رئيس جامعة اليرموك الموجه إلى معالي وزير التربية والتعليم، حيث قامت الوزارة بدورها بمخاطبة مدير إدارة الامتحانات والاختبارات التابعة لوزارة التربية والتعليم، والتي تقوم بإعداد وتطبيق الاختبار؛ إيذاناً بتسهيل المهمة لدراسة الأداء التفاضلي لاختباري العلوم والرياضيات وفقراتهما من "الاختبار الوطني لضبط نوعية التعليم" للصف العاشر تبعًا لطريقتي تقديم الاختبار(ورقي، محوسب)، وبالتحديد للحصول على المعلومات التالية:

أ- ما هو الاختبار؟

ب- مما يتكون؟

ج- ماذا يقيس؟

د- ما الغرض منه؟

ه- ما هي مجالات استخدام نتائجه؟

و- من يعده؟

ز- آلية تطبيقه؟

ح- آلية تصحيحه؟

ط- إجابات المفحوصين عن كل فقرة من فقرات اختباري العلوم والرياضيات عبر شكلي تقديمه، ممثلة بـــ (1) إذا أجاب المفحوص عن الفقرة إجابة صحيحة، و(0) إذا أجاب عنها إجابة خاطئة.

ي- محتوى الفقرات والمهارة التي تقيسها كل فقرة من الاختبار الذي قدم للمفحوصين خلال العام 2011/2010؟.

وعند مراجعة مديرية إدارة الامتحانات والاختبارات تبين، من خلال الزيارات واللقاءات الرسمية مع بعض المعنيين بالاختبار، ومن خلال التقرير السنوي الذي تعده الوزارة، أن الشكل المحوسب للاختبار يمثل صورة مماثلة للشكل الورقي من حيث: عدد الفقرات وترتيبها ونوعها، وعدد البدائل، وزمن الاختبار، إلا أنه على الشكل الورقي تظهر عدة فقرات على الصفحة الواحدة، بحيث يتراوح عددها ما بين (7-8) فقرات، بينما على الشكل المحوسب تظهر كل فقرة اختبارية على شاشة واحدة، مع إمكانية الرجوع للفقرات السابقة بالضغط على زر تأجيل السؤال، وإمكانية الانتقال إلى الفقرة التالية بالضغط على زر التالي، وإذا احتوى الاختبار على

نصوص فإنها تتكرر على كل شاشة، و ينتهي الاختبار عند انتهاء زمن الاختبار، أو إذا ضغط الطالب على زر (إنهاء)، حيث تظهر له عبارة (هل أنت متأكد: نعم، لا)، فإذا كانت الإجابة "نعم" يخرج الطالب من الاختبار، كما يمكن تكبير وتصغير الخط بالضغط على إشارتي (+ ، −) لمن يعاني من مشاكل في النظر، كما يوجد مؤقت للزمن يكون فيه التوقيت تنازلي؛ ليعرف الطالب ما تبقى من زمن الاختبار، بالإضافة إلى أن فقرات الشكل المحوسب للاختبار تكتب بحجم خط (16)، في حين تكتب فقرات الشكل الورقي بحجم خط (14)، بينما يكون نوع الخط (Simplified Arabic) على شكلي الاختبار.

كما تبين أن الاختبارات المحوسبة تطبق في المدارس في ضوء: عدد الطلبة، عدد مختبرات الحاسوب، عدد أجهزة الحاسوب المتصلة بشبكة الأنترنت، أو الأنترانت، وعرض الموجة (Band Width) بالنسبة لخط اتصال المدرسة بشبكة الأنترنت أو الأنترانت، كما أن كل طالب يحصل على نتيجته الكلية ونتيجته التفصيلية على كل فقرة حال انتهائه من تقديم الاختبار، وهذه النتيجة تنسخ مباشرة بشكل تلقائي إلى المركز الرئيسي للاختبارات، أما النسخة الورقية فيتم تصحيح إجابات الطلبة عنها بإدخال إجاباتهم إلى الحاسوب باستخدام القارئ الضوئي، وتصحيح الاختبار آليًا.

كما تم الحصول على البيانات اللازمة لإجابة أسئلة الدراسة، والتي تمثل إجابات طلبة الصف العاشر عن فقرات اختباري العلوم والرياضيات من الاختبار الوطني لضبط نوعية التعليم للصف العاشر بشكلين ورقي ومحوسب، على شكل إجابات ثنائية (0، 1)، وكانت البيانات على شكل ملف اكسل (Excel)، مرمزة حسب: الجنس، رقم الطالب، ورقم وجنس المدرسة، ورقم البديل الذي اختاره الطالب على كل فقرة من فقرات الاختبارين عبر شكلي الاختبار، بالإضافة إلى الدرجات الكلية للطلبة على الاختبارين.

تم تحويل هذه البيانات إلى ملف (SPSS)، بإدخال استجابات أفراد عينة الدراسة البالغ عددهم (2800) طالبًا وطالبة على اختباري العلوم والرياضيات المتكونان من (30) فقرة اختيار من متعدد على التوالي؛ لتسهيل عملية تحليل البيانات لغايات الكشف عن الأداء التفاضلي، كما تم الحصول على التقرير السنوي للاختبار بنسختين: ورقية، ومحوسبة على شكل ملف (Pdf) يحتوي على: تعريف الاختبار الوطني لضبط نوعية التعليم، أهداف الاختبار، مراحل تطور الاختبار، الطريقة والإجراءات لتحليل الاختبار من قبل مديرية الاختبارات والامتحانات، نتائج التحليل، والتوصيات.

بالإضافة إلى الحصول على نسخةٍ ورقيةٍ فقط من الاختبار برزمتيه، ومفتاح الإجابة لاختباري العلوم والرياضيات، ونسخ من نموذج القارئ الضوئي، ونظرًا لرفض قسم الاختبارات المحوسبة، لم يتم الاطلاع على فقرات الاختبار المحوسب، أو الحصول على نسخ ورقية منه للتعرف على مزايا فقراته وشكل إخراجه، لاعتبار ذلك يدخل في نطاق سرية الاختبار وخصوصية البيانات.

تم التحقق من الافتراضات الأساسية الواجب توفرها في البيانات عند تطبيق النموذج اللوجستي الثلاثي المعلمة ضمن نظرية الاستجابة للفقرة متمثلة بأحادية البعد، والاستقلال الموضعي وذلك على النحو الآتي:

افتراض أُحادية البعد(Unidimensionality Assumption)

تم التحقق من هذا الافتراض من خلال المؤشرات التي اعتمدت على التحليل العاملي (Factor analysis)؛ لتحديد البناء العاملي باستخدام طريقة للمكونات الرئيسية (**P**rincipal **C**omponents **A**nalysis, PCA)، وذلك لإجابات الأفراد عن فقرات اختباري العلوم والرياضيات بشكليهما؛ الورقي والمحوسب، من خلال برنامج (SPSS)، حيث يعتبر الاختبار أحادي البعد للسمة المقيسة، إذا كان نسبة ما يفسره العامل الأول كبيرة، وعادة ما تقدر بأكثر من

20% من التباين الكلي. ويمكن استخدام معيارًا أخر للحكم على أحادية البعد وهو أن تكون نسبة الجذر الكامن الأول إلى الجذر الكامن الثاني عالية، بحيث لا تقل عن 2، بالإضافة إلى استخدام التمثيل البياني (Scree Plot) لقيم الجذور الكامنة (Eigen Value) للعوامل المكونة للاختبار (Hambleton & Swaminathan, 1985; Hattie, 1985; Hulin, Drasgow & Parson, 1983)، وقد أجري التحليل العاملي لبيانات أدوات الدراسة، ويبين الجدول (7) قيم الجذور الكامنة، ونسبة التباين المفسر، وكذلك نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني، ونسبة الفرق بين الجذر الكامن للعامل الأول والجذر الكامن للعامل الثاني، إلى الفرق بين الجذر الكامن للعامل الثاني والجذر الكامن للعامل الثالث، وذلك لاختباري العلوم والرياضيات بشكليهما الورقي والمحوسب.

جدول (7)

التحليل العاملي لاستجابات المفحوصين على اختباري العلوم والرياضيات عبر شكلي تقديم الاختبار ورقي ومحوسب.

اختبار	طريقة تقديم الاختبار	المكون	الجذر الكامن	التباين المفسر	التباين المفسر التراكمي	الجذر الكامن الأول / الجذر الكامن الثاني	(الجذر الكامن الأول – الجذر الكامن الثاني) / (الجذر الكامن الثاني – الجذر الكامن الثالث)
رياضيات	ورقي	1	16.263	**54.209**	54.209	**10.10**	**39.28**
		2	1.610	5.365	59.575		
		3	1.237	4.122	**63.696**		
	حاسوبي	1	16.607	**55.357**	55.357	**9.48**	**29.29**
		2	1.752	5.840	61.196		
		3	1.245	4.149	**65.345**		
علوم	ورقي	1	15.397	**51.323**	51.323	**10.20**	**88.12**
		2	1.509	5.030	56.353		
		3	1.351	4.505	**60.857**		
	حاسوبي	1	15.296	**50.987**	50.987	**10.80**	**104.03**
		2	1.416	4.721	55.709		
		3	1.283	4.277	**59.985**		

يلاحظ من الجدول أن التحليل العاملي لبيانات اختبار الرياضيات بشكله الورقي قد أفرز ثلاثة عوامل الجذر الكامن لها يزيد على الواحد صحيح، فسر العامل الأول منها 54.209% من التباين، وفسرت جميع العوامل 63.696% من التباين الكلي، وبالتالي فإن نسبة التباين المفسر من العامل الأول هي نسبة أعلى من 20%، كما أن نسبة التباين المفسر لكل من العوامل المتبقية

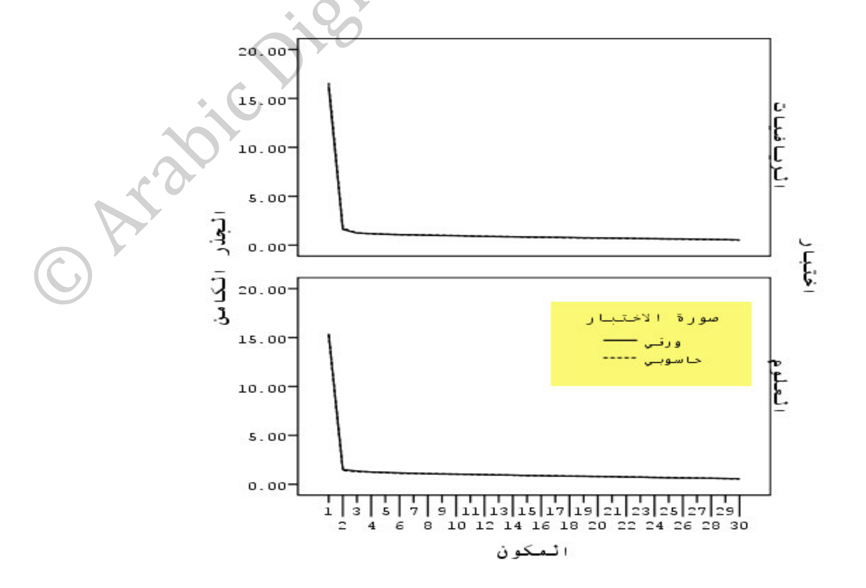
متقاربة، أي أن هناك تماثلا نسبيًا وشبه استقرار في نسب التباين المفسرة للعوامل جميعها باستثناء العامل الأول، وكانت قيمة الجذر الكامن الأول للعامل الأول 16.263 والثاني 1.610، وبالتالي فإن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني البالغة (10.10) أعلى من 2 وبناءً على ذلك يتبين تحقق أحادية البعد في بيانات هذا الاختبار.

كما يلاحظ من الجدول (7) أن التحليل العاملي لبيانات اختبار الرياضيات بشكله المحوسب قد أفرز ثلاثة عوامل الجذر الكامن لها يزيد على الواحد صحيح، فسر العامل الأول منها 55.357% من التباين، وبالتالي نسبة التباين المفسر من العامل الأول تكون أعلى من 20%، كما فسرت جميع العوامل 65.345 من التباين الكلي، في حين أن نسبة التباين المفسر لكل من العوامل المتبقية متقاربة، أي أن هناك تماثلا نسبيًا وشبه استقرار في نسب التباين المفسرة للعوامل جميعها باستثناء العامل الأول، وكانت قيمة الجذر الكامن الأول للعامل الأول 16.607 والثاني 1.509، وبالتالي فإن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني (9.48) أعلى من 2 وبناء على ذلك يتبين تحقق أحادية البعد في بيانات هذا الاختبار أيضاً.

أما بالنسبة لبيانات اختبار العلوم بشكله الورقي فقد أفرز التحليل العاملي أيضاً ثلاثة عوامل فسر العامل الأول منها 51.323 من التباين، حيث كانت النسبة أعلى من 20%، كما فسرت جميع العوامل 60.857 من التباين الكلي، حيث كانت نسبة التباين المفسر لكل من العوامل المتبقية متقاربة، وبالتالي هناك تماثلا نسبيًا وشبه استقرار في نسب التباين المفسرة للعوامل جميعها باستثناء العامل الأول، وكانت قيمة الجذر الكامن الأول للعامل الأول 15.397 والثاني 1.509، وبالتالي فإن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني (10.20) وهي أعلى من 2 وبناء على ذلك يتبين تحقق أحادية البعد في بيانات هذا الاختبار.

أما الشكل المحوسب من اختبار العلوم فقد أفرز التحليل العاملي ثلاثة عوامل فسر العامل الأول منها 50.987% من التباين، وفسرت جميع العوامل 59.985% من التباين الكلي، وكانت قيمة الجذر الكامن الأول للعامل الأول 15.296 والثاني 1.416، كما أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني (10.10) أعلى من 2 وبناءً على ذلك يتبين تحقق أحادية البعد في بيانات هذا الاختبار لأن نسبة التباين المفسر من العامل الأول هي نسبة أعلى من 20%، كما أن نسبة التباين المفسر لكل من العوامل المتبقية متقاربة، أي أن هناك تماثلا نسبيًا وشبه استقرار في نسب التباين المفسرة للعوامل جميعها باستثناء العامل الأول.

ويمكن تأكيد افتراض أحادية البعد لشكلي الاختبار عبر مجالي العلوم والرياضيات من خلال تمثيل الجذور الكامنة بيانيًا، من خلال ما يعرف بـــ (Scree Plot)، الذي يظهر بالشكل (3)، ويتبين من خلاله أن الجذر الكامن للعامل الأول يتميز بشكل واضح عن الجذور الكامنة لبقية العوامل، بحيث يمكن تسمية العامل الطاغي من بقية العوامل.



الشكل (3): رسم بياني لقيم الجذور الكامنة للعوامـــل المختلفـــة المتعلقـــة بالاســـتجابات لاختبـــاري العلـــوم والرياضيات بشكليهما الورقي والمحوسب.

كما تم التحقق من افتراض أحادية البعد باستخدام برنامج (**N**ormal **O**give **H**armonic **A**nalysis **R**obust **M**ethod, NOHARM) الذي أعده فريزر (Fraser, 1988)، حيث يقدم هذا البرنامج مؤشرين لفحص تحقق هذا الافتراض في البيانات:

- **مؤشر تاناكا Tanaka's Index**، هو مؤشر حسن مطابقة بين النموذج والبيانات، ويعمل كمعامل تحديد وتلخيص لنسبة التباين المفسر بواسطة النموذج المفترض ومعادلته هي:

$$Y_{ULS=1} - \frac{T_r(R_2)}{T_r(C_2)} ....(17)$$

حيث أن:

- R: مصفوفة التباين المشترك للبواقي.
- C: مصفوفة التباين المشترك للعينة.
- T_r: متتالية المصفوفة Trace of the Matrix

وقد أشار ماك دونالد إلى أن مؤشر تاناكا يعتبر دليلاً على المستوى المقبول من المطابقة بين النموذج والبيانات إذا بلغت قيمته 0.90، ويعتبر دليلاً على المستوى الجيد من المطابقة إذا بلغت قيمته 0.95، أما المطابقة التامة بين النموذج والبيانات فتحصل عندما تبلغ قيمة المؤشر إلى 1.00. (Jasper, 2010)

ويعرض الجدول(8) قيم مؤشر تاناكا في اختبارات الدراسة، حيث يتبين من الجدول أن مؤشر تاناكا للمجموعات المرجعية (الورقي) لاختباري العلوم والرياضيات قد تحقق فيها المستوى الجيد من المطابقة بين النموذج والبيانات، وأن المؤشر للمجموعات المستهدفة

(الحاسوبي) لاختباري العلوم والرياضيات قد تحقق فيها أيضًا المستوى الجيد من المطابقة بين النموذج والبيانات، وهذا دليل على تحقق افتراض أُحادية البعد في بيانات الدراسة.

- مؤشر الجذر التربيعي لمتوسطات مربعات البواقي Root Mean Square of Residuals(RMSR)

حيث يقوم برنامج (NOHARM) بإنتاج مصفوفة البواقي، بغرض إجراء تحليلات مطابقة النموذج للبيانات، وتكون قيم هذه المصفوفة هي الفرق بين التباينات المشتركة الملاحظة و التباينات المشتركة الناتجة من إجراء مطابقة النموذج للبيانات، وتكون المطابقة تامة إذا كان الفرق بينهما يساوي صفر.

ثم يقوم البرنامج بتلخيص مصفوفة البواقي من خلال إنتاج الجذر التربيعي لمتوسط المربعات RMSR، وهو الجذر التربيعي لمعدل الفرق المربع بين التباينات المشتركة الملاحظة والتباينات المشتركة المتنبأ بها، وهكذا فإن القيم المنخفضة للجذر التربيعي لمتوسط مربعات البواقي مؤشر على حسن المطابقة، ومن أجل تقدير قيمة هذا المؤشر فقد اقترح فريزر(Fraser, 1988) مقارنة قيمة المؤشر مع قيمة الخطأ المعياري للبواقي كمعيار والذي يتم حسابه وفق المعادلة التالية:

$$RMSR \leq 4.1\sqrt{\mathrm{N}}...............(18)$$

حيث N:حجم العينة.

ويعرض الجدول(8) قيم مؤشر الجذر التّربيعي لمتوسطات مربعات البواقي.

جدول (8)

قيم مؤشر الجذر التربيعي لمتوسطات مربعات البواقي، وقيم مؤشر تاناكا على اختباري العلوم والرياضيات عبر شكلي تقديم الاختبار ورقي ومحوسب.

الاختبار	إحصائيات أحادية البعد	ورقي		حاسوبي	
		القيمة	نقطة القطع لاحتمالية قبول قيمة الإحصائي	القيمة	نقطة القطع لاحتمالية قبول قيمة الإحصائي
رياضيات	مجموع مربعات البواقي (دون القطر الرئيسي)	0.04155472		0.04716837	
	الجذر التربيعي لمتوسط مجموع مربعات البواقي (دون القطر الرئيسي)	0.01227031	**0.15619756**	0.01307287	**0.156084332**
	مؤشر Tanaka لحسن المطابقة	0.95831561		0.94744481	
علوم	مجموع مربعات البواقي (دون القطر الرئيسي)	0.03652296		0.04461315	
	الجذر التربيعي لمتوسط مجموع مربعات البواقي (دون القطر الرئيسي)	0.00982963	**0.156084332**	0.01086389	**0.155521865**
	مؤشر Tanaka لحسن المطابقة	0.94877701		0.93517859	

ويتبين من الجدول (8) أن جميع قيم المؤشر كانت منخفضة جدًا وتقترب من الصفر، وكانت أقل من المعيار، وهذا دليل كاف على تحقق افتراض أحادية البعد في بيانات الدراسة.

- **افتراض الاستقلال الموضعي (Local Independence Assumption)**

يعد تحقق افتراض الاستقلال الموضعي نتيجة محصلة من تحقق افتراض أحادية البعد (Hulin, Drasgow & Parsons, 1983)، في حين ينظر إليه بعض الباحثين على أنه افتراض مكافئ لافتراض أحادية البعد (Hambleton & Swaminathan, 1985)، وبالتالي إذا تم التحقق من افتراض أحادية البعد، فإن افتراض الاستقلال الموضعي يتحقق أيضاً، في حين اعتبر آخرون أن الاستقلال الموضعي لا يكافئ أحادية البعد (Meara, Robin & Sireci, 2000)، لذلك، ولضمان توفر الاستقلال الموضعي، تم فحص تحقق هذا الافتراض في البيانات؛ بحساب مؤشر Z_{Q3}، ويعرف مؤشر Q3 بأنه: معامل ارتباط بيرسون للبواقي بين زوج من الفقرات بعد ضبط السمة المقدرة، حيث عمل شن (Shen, 1997) على تعديل هذا المؤشر، وذلك بتحويل الأخطاء الملاحظة إلى معيارية، بحيث يتم حساب هذا المؤشر وفق المعادلة التالية:

$$Z_{Q3} = 1/2Log\,[\frac{1+\text{Q3}}{1\text{-}\text{Q3}}]..............(19)$$

وللحكم على تحقق الاستقلال الموضعي ينبغي أن تقع قيمة Z_{Q3} ضمن انحرافين معياريين عن الوسط الحسابي لكافة قيم Z_{Q3}، ويحسب الانحراف المعياري للمؤشر حسب المعادلة:

$$\sigma_{Z_{Q3}} = \sqrt{1/N-3}.............(20)$$

ولحساب هذا المؤشر استخدم برنامج ((**L**ocal **D**ependence **I**ndices for **D**ichotomous Items :LDID)) الذي أعده كيم وكوهين (Kim & Cohen, 2005) وتم حساب عدد أزواج الفقرات والأوساط الحسابية والانحرافات المعيارية لقيم Z_{Q3} وفترات الثقة للمجموعتين المرجعية والمستهدفة في كل اختبار كما هو موضح في الجدول(9):

جدول (9)

أزواج الفقرات والمتوسطات الحسابية والانحرافات المعيارية لقيم Z_{Q3}

الاختبار	الإحصائي		ورقي	حاسوبي
رياضيات	عدد الأزواج		276	276.000000000000
	مدى القيم المشاهدة	القيمة الدنيا	-0.147490000000	-0.161930000000
		القيمة العظمى	0.182010000000	0.237390000000
	المتوسط الحسابي المشاهد		**-0.009186090000**	**-0.008459021739**
	الانحراف المعياري المشاهد		0.057425266000	0.067566965577
	المتوسط الحسابي النظري		**-0.043478260870**	**-0.043478260870**
	التباين لـ Zq3		0.001457725948	0.001455604076
	الانحراف المعياري لـ Zq3		0.038180177416	0.038152379686
	الانحراف المعياري لـ 2× Zq3		0.076360354832	0.076304759372
	فترة الثقة	الحد الأدنى	-0.0855464448321212	-0.0847637811107820
		الحد الأعلى	0.0671742648321213	0.0678457376325211
علوم	عدد الأزواج		378	378
	مدى القيم المشاهدة	القيمة الدنيا	-0.258510000000	-0.409720000000
		القيمة العظمى	0.255890000000	0.240630000000
	المتوسط الحسابي المشاهد		**-0.020398253968**	**-0.019692301587**
	الانحراف المعياري المشاهد		0.065422903870	0.068471681558
	المتوسط الحسابي النظري		**-0.037037037037**	**-0.037037037037**
	التباين لـ Zq3		0.001455604076	0.001445086705
	الانحراف المعياري لـ Zq3		0.038152379686	0.038014296063
	الانحراف المعياري لـ 2× Zq3		0.076304759372	0.076028592127
	فترة الثقة	الحد الأدنى	-0.0967030133399056	-0.0957208937142721
		الحد الأعلى	0.0559065054033976	0.0563362905396690

كما تم حساب عدد أزواج الفقرات التي وقعت ضمن فترة الثقة المحققة لشرط الاستقلال الموضعي لطلبة كلا المجموعتين (المرجعية ورقي، المستهدفة حاسوبي) على اختباري العلوم والرياضيات، والجدول 10 يبين مؤشرات الاستقلال الموضعي وفقاً لنظرية استجابة الفقرة.

جدول (10)

مؤشرات الاستقلال الموضعي وفقاً لنظرية استجابة الفقرة

الاختبار	حالة الاستقلال الموضعي	ورقي		حاسوبي	
		عدد الأزواج	النسبة المئوية	عدد الأزواج	النسبة المئوية
رياضيات	معتمد	44	15.9	61	22.1
	مستقل	232	84.1	215	77.9
	الكلي	**276**	**100.0**	**276**	**100.0**
علوم	معتمد	74	19.6	80	21.2
	مستقل	304	80.4	298	78.8
	الكلي	**378**	**100.0**	**378**	**100.0**

يلاحظ من الجدول 10، أن النتائج الخاصة به قد كانت على النحو الآتي:

- فيما يخص الشكل الورقي لاختبار الرياضيات: عدد أزواج الفقرات التي وقعت خارج المدى (44) بنسبة (15.9%)، بينما كان عدد أزواج الفقرات التي وقعت ضمن المدى (232) بنسبة (84.1%)، وهذا يبين أن عدد أزواج الفقرات التي حققت الاستقلالية أعلى من عدد أزواج الفقرات التي حققت التبعية الموضعية، وهذا مؤشر على تحقق افتراض الاستقلال الموضعي.

- فيما يخص الشكل المحوسب لاختبار الرياضيات: عدد أزواج الفقرات التي وقعت خارج المدى (61) بنسبة (22.1%)، بينما كان عدد أزواج الفقرات التي وقعت ضمن المدى (215) بنسبة (77.9%)، وهذا يبين أن عدد أزواج الفقرات التي حققت الاستقلالية أعلى من عدد أزواج الفقرات التي حققت التبعية الموضعية، وهذا مؤشر على تحقق افتراض الاستقلال الموضعي.

- فيما يخص الشكل الورقي لاختبار العلوم: عدد أزواج الفقرات التي وقعت خارج المدى (74) بنسبة (19.6%)، بينما كان عدد أزواج الفقرات التي وقعت ضمن المدى (304) بنسبة (80.4%)، وهذا يبين أن عدد أزواج الفقرات التي حققت الاستقلالية أعلى من عدد أزواج الفقرات التي حققت التبعية الموضعية، وهذا مؤشر على تحقق افتراض الاستقلال الموضعي.
- فيما يخص الشكل المحوسب لاختبار العلوم: عدد أزواج الفقرات التي وقعت خارج المدى (80) بنسبة (21.2%)، بينما كان عدد أزواج الفقرات التي وقعت ضمن المدى (298) بنسبة (78.8%)، وهذا يبين أن عدد أزواج الفقرات التي حققت الاستقلالية أعلى من عدد أزواج الفقرات التي حققت التبعية الموضعية، وهذا مؤشر على تحقق افتراض الاستقلال الموضعي.

ومن الجدير بالذكر أن افتراضات نظرية الاستجابة للفقرة لن تتحقق بشكل مطلق في أي مجموعة من البيانات (Albanese & Foresyth, 1984).

التوصل إلى عينة الدراسة

إن البيانات التي تم الحصول عليها بيانات حقيقية لإجابات المفحوصين عن فقرات الاختبار الوطني لضبط نوعية التعليم للصف العاشر؛ بهدف تحري الأداء التفاضلي على فقرات اختباري العلوم والرياضيات المستمدين من الاختبار تبعا لطريقة تقديم الاختبار، بغض النظر عن كونه مقنن أولا؛ لكنّ برمجيات التحليل المستخدمة لتحقيق هدف الدراسة تضع محددات تتعلق بمطابقة البيانات للنموذج المستخدم في الدراسة، فتطبيق النموذج الثلاثي المعلمة يتطلب

أحجام عينات كبيرة يصل حجمها (N>1000) (Crocker & Algina, 1986). كما أن الإحصائي اللامعلمي χ^2 المستخدم، لغايات المطابقة حساس لأحجام العينات الكبيرة

وبالرغم من الدقة التي توفرها طريقة "الكشف عن الأداء التفاضلي للفقرة والاختبار" المعتمدة على مؤشرات الأداء التفاضلي غير التعويضي للفقرة، والأداء التفاضلي التعويضي للفقرة، من خلال ارتباطها بالدرجة الكلية على الاختبار، إلا أن طبيعة الإجراءات التي يتطلبها الكشف عن الأداء التفاضلي بهذه الطريقة قائمة على تحقق افتراضات النموذج المستخدم، وعلى التقدير الدقيق لمعالم الفقرات، والتقدير الدقيق لمعالم الفقرات يتطلب أحجام عينات كبيرة؛ وسعيًا إلى أن تكون العينة التي تم اختيارها متقاربة مع ما ذكر، ولضمان أن تأتي النتائج في خدمة أغراض الدراسة، تم اختيار عينة الدراسة من المجتمع، بحيث تكون مطابقة للنموذج المستخدم، حيث تم التعامل مع ملفات البيانات على ما هي عليه، من حيث حجومها (رياضيات ورقي 49073، رياضيات حاسوبي 6064، علوم ورقي 45227، علوم حاسوبي 6076)، إلاَّ أنه وبسبب عدم مطابقة كافة فقرات كلٍّ من الاختبارين على اختلاف صورتيهما بحجومها كما وردت من المصدر، تم التعامل معها باختيار عينات عشوائية من كل ملف بيانات بما يتناسب مع حجم البيانات الخاصة بكل منها، حيث تم أخذ عينات عشوائية من البيانات كبيرة الحجم التي تغطي اختباري الرياضيات والعلوم بصورتهما الورقية ابتدأت بعشر آلاف، ونظراً لعدم مطابقة أيٍّ من فقراتها، تم أخذ عينات عشوائية ابتدأت بخمسة آلاف، ونظراً لعدم مطابقة أيٍّ من فقراتها تم أخذ عينات عشوائية ابتدأت بألف في كل مرة.

وفي الجهة المقابلة في البيانات صغيرة الحجم نسبياً التي تغطي اختباري الرياضيات والعلوم بصورتهما الحاسوبية فقد تم أخذ عينات عشوائية ابتدأت بخمسة آلاف، ثم أربعة آلاف،

ثم ثلاثة آلاف، ثم ألفين ثم ألف، وفي المحصلة، تم التوفيق بين أحجام العينات المختارة لاختباري الرياضيات والعلوم بصورتهما الورقية وبين اختباري الرياضيات والعلوم بصورتهما الحاسوبية إلى أن وصل حجم العينات لصورتي الاختبارين الورقي والحاسوبي 700 طالباً وطالبة لكلٍ منهما، حتى تم اعتماد عينة مقدارها (2,800) طالباً وطالبة بصورة عشوائية من مجتمع الدراسة، حيث تم تقسيم العينة إلى مجموعتين المجموعة المرجعية وتتكون من (1400) مفحوصاً لكلٍ من الاختبارين تمثل الشكل الورقي للاختبارين وذلك لأسباب منطقية كون الشكل الورقي هو الشكل التقليدي الشائع الاستخدام في عملية التقويم، ويمثل المعيار الذي يتم مقارنة المجموعة المستهدفة عليه، وللتعرف على مدى استفادة الطلبة من مزايا الاختبارات المحوسبة على الشكل المحوسب للاختبار.

بينما تتكون المجموعة المستهدفة من (1400) طالباً وطالبة لكلٍ من الاختبارين وتمثل الشكل المحوسب للاختبار، والذي يعتقد بأن الفقرات تتحيز ضدها، كما تم تقسيم كل مجموعة من المجموعتين إلى مجموعتين فرعيتين (علوم ورياضيات) متساويتين في العدد ومتناظرتين بالقدرة؛ أي بالمحصلة هناك حاجه لوجود أربع مجموعات من الأفراد لتخضع للتحليل بالاعتماد على طريقة تقديم الاختبار (علوم ورقي، علوم محوسب، رياضيات ورقي، رياضيات محوسب)، من أجل تحري الأداء التفاضلي بينها على فقرات اختبار ضبط نوعية التعليم للصف العاشر للعام الدراسي 2011/2010

المطابقة الإحصائية

المطابقة للأفراد والفقرات:

يعد فحص حسن المطابقة للبيانات حسب النموذج اللوجستي الثلاثي المعلمة ضمن نظرية الاستجابة للفقرة من الإجراءات المهمة؛ كون مميزات نماذج نظرية الاستجابة للفقرة التي تتعلق بتفسير النتائج يمكن أن تتحقق إذا كان هناك حسن مطابقة بين النموذج المستخدم والبيانات، حيث تمّ تحليل البيانات الخام باستخدام برمجية (BILOG-MG 3)، لمطابقة الأفراد للنموذج اللوجستي الثلاثي المعلمة، حيث توفر هذه البرمجية تقديرات لمعالم الفقرة الثلاث: الصعوبة، التمييز، التخمين، وتقديرات القدرة للمفحوصين. وتستخدم هذه البرمجية الإحصائي χ^2 للمطابقة (χ^2 Statistic Fit) عند مستوى الدّلالة (α=0.01) للتدليل على مدى مطابقة النموذج المستخدم للبيانات.

وقد مرت إجراءات المطابقة للأفراد والفقرات بثلاث مراحل بإدخال إجابات المفحوصين والبالغ عددهم (2800) طالب وطالبة عن فقرات اختباري العلوم والرياضيات عبر شكلي تقديم الاختبارين (ورقي ومحوسب)، إذ أُدخلت إجابات الطلبة عن كل منها على حدة، تم إجراء تحليل البيانات للمرة الأولى، وذلك لحذف الأفراد غير المطابقين للنموذج اللوجستي ثلاثي المعلمة، بحيث يتم حذف الأفراد غير المطابقين للنموذج إذا كانت قيمة الاحتمالية لهم أقل من (α=0.01)، وقد أفرزت نتائج التحليل عدم مطابقة استجابات مجموعة من المفحوصين، كانت قيمة الاحتمالية لهم أقل من (0.01)، ويبين الجدول (11) أعداد الطلبة وأرقام المفحوصين غير المطابقين للنموذج اللوجستي ثلاثي المعلمة.

جدول (11)

الأفراد غير المطابقين في كلٍ من الاختبارين على اختلاف صورتي الاختبار

الاختبار	صورة الاختبار	رقم الطالب	عدد الإجابات الصحيحة	نسبة الإجابة الصحيحة	القدرة	الخطأ المعياري	احتمالية الخطأ
رياضيات	ورقي	6615	3	12.50%	-4.000	999	1.000
		7790	2	8.33%	-4.000	999	1.000
		10671	4	16.67%	-4.000	999	1.000
		22100	4	16.67%	-4.000	999	1.000
		22497	3	12.50%	-4.000	999	1.000
		22657	1	4.17%	-4.000	999	1.000
		25457	3	12.50%	-4.000	999	1.000
		47744	2	8.33%	-4.000	999	1.000
		49001	3	12.50%	-4.000	999	1.000
		49013	2	8.33%	-4.000	999	1.000
		49068	7	29.17%	-4.000	999	1.000
	حاسوبي	51212	5	20.83%	-4.000	999	1.000
		51362	1	4.17%	-4.000	999	1.000
		51683	2	8.33%	-4.000	999	1.000
		52530	2	8.33%	-4.000	999	1.000
		53378	2	8.33%	-4.000	999	1.000
		53465	1	4.17%	-4.000	999	1.000
		54214	1	4.17%	-4.000	999	1.000
		54439	2	8.33%	-4.000	999	1.000
		54761	3	12.50%	-4.000	999	1.000
		54925	3	12.50%	-4.000	999	1.000
علوم	ورقي	4194	6	24.00%	-4.000	999	1.000
		6846	5	20.00%	-4.000	999	1.000
		11790	6	24.00%	-4.000	999	1.000
		13364	14	56.00%	-0.356	0.662	0.004
		13449	4	16.00%	-4.000	999	1.000
		13589	5	20.00%	-4.000	999	1.000
		24769	6	24.00%	-4.000	999	1.000
		26525	5	20.00%	-4.000	999	1.000
		35812	6	24.00%	-4.000	999	1.000
		40841	6	24.00%	-4.000	999	1.000
	حاسوبي	45516	8	32.00%	-4.000	999	0.253
		49193	6	24.00%	-4.000	999	1.000
		49764	5	20.00%	-4.000	999	1.000
		50327	4	16.00%	-4.000	999	1.000
		50768	18	72.00%	0.767	0.348	0.009

بعد ذلك تم حذف استجابات الأفراد التي لم تتطابق استجاباتهم مع توقعات النموذج، حيث

بلغ عدد الأفراد غير المطابقين للنموذج (36) مفحوصاً منهم (21) مفحوصًا على اختبار

الرياضيات و(15) مفحوصًا على اختبار العلوم لم تتطابق استجاباتهم الملاحظة مع الاستجابات المتوقعة تبعاً لقدراتهم؛ أي أنها تبتعد عن توقعات النموذج.

بعد استبعاد الأفراد غير المطابقين للنموذج اللوجستي ثلاثي المعلمة تم إعادة التحليل مرة أخرى؛ لمعرفة مدى مطابقة الفقرات للنموذج، بحيث يتم حذف الفقرات غير المطابقة للنموذج إذا كانت قيمة الاحتمالية لهم أقل من (α=0.01)، وتبين عدم مطابقة بعض الفقرات للنموذج اللوجستي ثلاثي المعلمة، إذ كانت قيمة الاحتمالية لها أقل من (0.01)، ويبين الجدول (12) أرقام الفقرات غير المطابقة للنموذج اللوجستي ثلاثي المعلمة حسب تسلسل أرقامها في نماذج الاختبار في مجموعات الدراسة التي أشارت نتائج التحليل أنها غير متطابقة مع توقعات النموذج، وكانت الاستجابات الملاحظة عنها تبتعد عن الاستجابات التي يتوقعها النموذج.

الجدول (12)

الفقرات غير المطابقة في كلٍّ من اختباري العلوم والرياضيات على اختلاف صورتي الاختبارين

الاختبار	صورة الاختبار	رقم الفقرة	معلمة التمييز		معلمة الصعوبة		معلمة التخمين		كا2 لحسن المطابقة	درجة الحرية	مطابقة الفقرة
			قيمتها	الخطأ المعياري لها	قيمتها	الخطأ المعياري لها	قيمتها	الخطأ المعياري لها			
رياضيات	ورقي	11	1.90	0.22	-0.03	0.09	0.13	0.04	**36.50**	7	0.000
		18	1.94	0.22	-0.35	0.10	0.14	0.05	**51.00**	6	0.000
	حاسوبي	7	2.26	0.30	-0.34	0.12	0.28	0.05	**21.90**	6	0.001
		11	1.68	0.18	-0.27	0.11	0.15	0.05	**42.70**	7	0.000
		17	6.36	2.00	1.41	0.06	0.34	0.02	**36.30**	8	0.000
		18	1.93	0.24	-0.38	0.13	0.24	0.06	**29.00**	6	0.000
		22	2.02	0.28	-0.15	0.13	0.28	0.05	**20.80**	6	0.002
		30	5.29	1.37	1.71	0.08	0.15	0.02	**24.70**	8	0.002
علوم	ورقي	6	1.42	0.19	-0.10	0.15	0.18	0.06	**25.20**	7	0.001
		8	1.39	0.19	-0.30	0.17	0.20	0.06	**27.20**	8	0.001
		16									كلاسيك
		17									كلاسيك
		21	1.83	0.25	-0.29	0.14	0.25	0.06	**25.80**	7	0.001
	حاسوبي	6	1.73	0.23	0.04	0.12	0.19	0.05	**30.00**	7	0.000
		8	1.46	0.20	-0.48	0.19	0.25	0.07	**18.00**	6	0.006
		16									كلاسيك
		17									كلاسيك
		21	2.03	0.27	-0.53	0.15	0.28	0.07	**36.50**	5	0.000

يلاحظ من الجدول السابق، وجود (5) فقرات من اختبار العلوم على كلا شكلي الاختبار الورقي والمحوسب غير مطابقة للنموذج اللوجستي ثلاثي المعلمة، حيث حذفت الفقرات (6, 8, 16, 17, 21) من كلا شكلي الاختبار الورقي والمحوسب من بينها الفقرتين (16, 17) حذفت في التشغيل الأول لأسباب لها علاقة بالنظرية الكلاسيكية على كلا شكلي الاختبار، وحذفت الفقرات (6, 8, 21) في التشغيل الثاني لبرنامج BILOG-MG 3 لأسباب لها علاقة بنظرية الاستجابة للفقرة، من حيث انخفاض احتمالية الخطأ لهذه الفقرات دون مستوى الدلالة المتبنى في برنامج BILOG -MG 3 البالغة قيمته (0.01)

أما بالنسبة لاختبار الرياضيات الورقي فقد كان هناك فقرتين غير مطابقتين للنموذج اللوجستي ثلاثي المعلمة، حيث تم حذف الفقرتين (11, 18) في التشغيل الثاني لبرنامج BILOG -MG 3 لأسباب لها علاقة بنظرية استجابة الفقرة من حيث انخفاض احتمالية الخطأ لهاتين الفقرتين دون مستوى الدلالة المتبنى في برنامج BILOG -MG 3 البالغة قيمته (0.01)، كما كان هناك (6) فقرات من اختبار الرياضيات المحوسب غير مطابقة للنموذج اللوجستي ثلاثي المعلمة، حيث حذفت الفقرات (7, 11, 17, 18, 22, 30) لأسباب لها علاقة بنظرية استجابة الفقرة، من حيث انخفاض احتمالية الخطأ لهذه الفقرات دون مستوى الدلالة المتبنى في برنامج BILOG -MG 3 البالغة قيمته (0.01) وبالتالي تحذف الفقرتين (11, 18) من كلا شكلي الاختبار الورقي والمحوسب. وأخيراً تم إسقاط الفقرات (7, 17, 22, 30) من الشكل الورقي للاختبار ليكون هناك اتساق مع إسقاطها من الشكل المحوسب للاختبار، للتوفيق بين عدد الفقرات عبر شكلي الاختبار؛ بهدف تأمين نفس العدد من الفقرات المتقابلة، من حيث الرقم الخاص بكل واحدة منها للاختبار عبر شكلية تمهيدا لاستخدامها لغايات الكشف عن الأداء التفاضلي.

تم إعادة التحليل للمرة الثالثة بعد حذف الفقرات غير المطابقة للحصول على تقديرات نهائية لكل من معالم الفقرات متحررة من قدرات الأفراد، وقدرات الأفراد متحررة من معالم الفقرات على النموذج اللوجستي ثلاثي المعلمة. (Zimowski et al., 1996)

وبذلك تم الحصول على أدوات الدراسة: اختبار العلوم الذي يتكون من (25) فقرة عبر شكلي تقديمه بنسبة (83.33%) تقريبًا، واختبار الرياضيات الذي يتكون من (24) فقرة بنسبة (80%) من فقرات الاختبار الكلي عبر شكلي تقديمه، وكانت هذه الفقرات ممثلة للنطاق السلوكي للسمة تمثيلاً متوازناً، حسب عناصر المحتوى، ومستويات نتاجات التعلم المعرفية، وبناءً عليه تم اعتماد عينة المفحوصين المتبقية البالغ عددها (2764) طالباً وطالبة من كلا شكلي تقديم الاختبار الورقي والمحوسب منهم (1379) طالباً وطالبةً عبر الشكل الورقي للاختبار، و(1385) طالب وطالبة عبر الشكل المحوسب للاختبار، والتي تم استخدامها كأساس في إجراء التحليلات الإحصائية المختلفة للكشف عن الأداء التفاضلي تمهيدا للإجابة على أسئلة الدراسة.

المعالجات الإحصائية:

- **مطابقة الأفراد للنموذج اللوجستي الثلاثي المعلمة في نظرية الاستجابة للفقرة.**

تم استخدام برمجية (BILOG -MG 3) لمطابقة الأفراد للنموذج اللوجستي الثلاثي المعلمة، باستخدام اختبار χ^2 عند مستوى الدلالة (α=0.01).

- **مطابقة الفقرات للنموذج اللوجستي ثلاثي المعلم في نظرية الاستجابة للفقرة.**

تم استخدام برمجية (BILOG -MG 3)، لمطابقة الفقرات للنموذج اللوجستي الثلاثي المعلمة، باستخدام اختبار χ^2 عند مستوى الدلالة (α=0.01).

- **تقدير معالم الفقرة (الصعوبة، التمييز، والتخمين) ومعالم القدرة.**

باستخدام برنامج (BILOG-MG 3) من خلال استجابات المفحوصين لفقرات مجموعتي الدراسة (المرجعية والمستهدفة) لكل مجموعة على حدةٍ باستخدام طريقة الأرجحية العظمى الهامشية (**M**arginal **M**aximum **L**ikelihood, MML) التي تعتبر من أنسب الطرق في تقدير المعالم، وتعمل هذه الطريقة على إيجاد قيمة (θ) التي تجعل Log^{θ} أكبر ما يمكن من خلال الصيغة الرياضية التالية:

$$\log L_i(\theta) = \sum_{i=1}^{n}\{X_{ij}\,\text{loge}\,\ddot{e}_i(\theta) + (1 - X_{ij})\,\text{loge}[1 - P_j(\theta)]\} \ldots (21)$$

حيث أن: $P_j(\theta)$ دالة الاستجابة للفقرة j.

ونحصل على أكبر قيمة لــ (θ) عند مساواة المشتقة الأولى للمعادلة (21) بالنسبة لها، ومساواة المعادلة بالصفر، وإيجاد حلها.

هذا وقد تم استخراج نتائج التحليل وفق النموذج الثلاثي المعلمة وذلك لشكلي الاختبار وفق المراحل الثلاث الآتية:

- المرحلة الأولى (Phase 1): تعطي نتائج التحليل الخاصة بالفقرة والاختبار وفق النظرية الكلاسيكية في القياس، حيث تعرض مخرجات هذه المرحلة معامل الصعوبة للفقرة معرفة على أنها نسبة الطلبة الذين أجابوا إجابة صحيحة من بين المفحوصين الذين حاولوا الإجابة عن هذه الفقرات، وتقدير أخر لصعوبة الفقرة باللوجيت، ومعامل التمييز للفقرة (Item*Test Correlation) معرف بمعامل ارتباط بيرسون (Bearson)، ومعامل ارتباط الثنائي النقطي بايسيريال (Bi-serial).

- المرحلة الثانية (Phase 2): تعطي نتائج التحليل الخاصة بمعالم الفقرات وفق نظرية الاستجابة للفقرة والأخطاء المعيارية لكل معلم (SE)، وإحصائي حسن المطابقة لكل فقرة (اختبار χ^2 عند مستوى الدلالة α=0.01)، ومعالم الفقرات وهي: الصعوبة (Threshold)، والتمييز (Slope)، والتخمين (Asymptote).

- المرحلة الثالثة (Phase 3): تعطي نتائج التحليل الخاصة بتقديرات قدرات الأفراد (Ability)، وقيمة الخطأ المعياري لكل قدرة، وإحصائي حسن المطابقة لكل فرد (اختبار χ^2 عند مستوى الدلالة α=0.01) وعلامات الأفراد الخام على الاختبار ونسبتهم المئوية.

- **المعالجات الإحصائية المتعلقة بالكشف عن الأداء التفاضلي:**

- **المعادلة لشكلي الاختبار عبر مجالي العلوم والرياضيات:**

وضع جميع تقديرات المعالم على مقياس مشترك (Common Metric) باستخدام برنامج (Equate v2.1) (Baker, 1993) بجعل المتوسط الحسابي للقدرة يساوي صفر، والانحراف المعياري لها يساوي واحد صحيح، وذلك بوضع معالم الفقرة الصعوبة b والتمييز a والتخمين c وتبايناتها Variances وتبايناتها المصاحبة Covariances المقدرة من المجموعة المستهدفة (الشكل المحوسب) على مقياس المجموعة المرجعية (الشكل الورقي)، حيث يعمل التحويل الخطي على التقليل من مجموع مربعات الفرق بين معالم الفقرة المتقابلة عبر المجموعتين مما يجعل النتائج قابلة للمقارنة.

وكون عملية المعادلة تتطلب وجود مجموعة من الفقرات المشتركة (Common Anchor Items) خالية من الأداء التفاضلي لحساب معاملات المعادلة، ونظرا لعدم تحقق هذا

المحدد قبل الكشف عن الأداء التفاضلي؛ فإنه تم التغلب على هذه المشكلة بإجراء المعادلة على مرحلتين 2-Stage Linking (Candell & Drasgow, 1988)، وذلك وفق الخطوات التالية:

- استخدام جميع فقرات الاختبار كفقرات مشتركة في المرحلة الأولية من المعادلة لحساب معاملات المعادلة الثابت الضربي (Multiplicative Constant) والثابت الجمعي (Additive Constant) باستخدام برنامج (Equate v2.1).
- تحديد الفقرات التي تبدي أداءً تفاضليًا كبيرًا من خلال استخدام برنامج (Dfit v8.04) (Raju, Oshima, & Wolach, 2009).
- استثناء الفقرات التي أبدت أداءً تفاضليًا كبيرًا من الفقرات المشتركة، وليس من جسم الاختبار.
- إجراء المعادلة للمرة الثانية باستخدام الفقرات المشتركة المتبقية التي لم تظهر أداءً تفاضليًا من المرحلة الأولى باستخدام برنامج (Equate v2.1) مرة أخرى، للوصول إلى معاملات المعادلة (الثابت الضربي والثابت الجمعي).
- استخدام معاملات المعادلة (الثابت الضربي والثابت الجمعي) التي يتم حسابها من المرحلة الثانية للمعادلة لحساب مؤشرات (DFIT) النهائية.

- **الكشف عن الأداء التفاضلي:**

تم الكشف عن الأداء التفاضلي في فقرات اختباري العلوم والرياضيات تبعاً لشكل تقديمهما (ورقي، محوسب) باستخدام إجراءات طريقة "الأداء التفاضلي للفقرة والاختبار (DFIT)" في النموذج اللوجستي ثلاثي المعلمة، بحساب قيم مؤشر "الأداء التفاضلي التعويضي للفقرة

(CDIF)"، حسب المعادلة (15) لجميع فقرات الاختبار، ثم جمع هذه القيم للحصول على مؤشر "الأداء التفاضلي للاختبار DIF" حسب المعادلة (14) الواردة في الفصل الأول من الدراسة. كما تم حساب مؤشر "الأداء التفاضلي غير التعويضي للفقرة (NCDIF)" لكل فقرة من فقرات الاختبار على حدة تحت افتراض خلو فقرات الاختبار الأخرى من الأداء التفاضلي كما في المعادلة (9) الواردة في الفصل الأول من الدراسة.

هذا وتم فحص الدلالة الإحصائية لمؤشرات الأداء التفاضلي غير التعويضي للفقرة، والأداء التفاضلي للاختبار باستخدام طريقة "معلمة الفقرة التكرارية (IPR)" التي تمثل طريقة لاشتقاق درجة قطع معينة لكل فقره من فقرات الاختبار، حيث تعتبر درجة القطع المحددة معيارًا للحكم على وجود الأداء التفاضلي للفقرة أو عدم وجوده، عندما تكون قيمة المؤشر المحسوب أعلى من درجة القطع، وتحسب درجة القطع عن طريق تقدير معالم الفقرة، وحساب التباين ((V(c), V(a), V(b)) والتباين المشترك ((Cov(a,c), Cov(b,c), Cov(b,a)) لهذه التقديرات، وتكرار ذلك عدد كبير من المرات تصل في حدها الأقصى (1,000 تكرار للمجموعة المستهدفة)، للحصول على عدد كبير من قيم (NCDIF)، وترتيبها وإعطائها رتب مئينيه، تمثل توزيع تكراري لهذه القيم، وتحدد درجة القطع لمؤشر الأداء التفاضلي غير التعويضي للفقرة عند قيمة NCDIF المرتبطة برتبة مئينية من التوزيع التكراري لقيم (NCDIF) عند مستويات دلالة مختلفة تتراوح بين (0.001 إلى 0.5).

أما درجة القطع لمؤشر الأداء التفاضلي للاختبار فهي عبارة عن مجموع درجات القطع لمؤشر NCDIF لجميع فقرات الاختبار، أما الدلالة الإحصائية لمؤشر الأداء التفاضلي التعويضي فيتم فحصه عن طريق "مؤشر إزالة الفقرة 2CDIF– Item removal index)

(NCDIF" الذي يقوم بحذف الفقرات ذات قيمة (CDIF) العالية مرة بعد مــرة حتــى تصــبح قيمةDTF غير دالة إحصائياً، بحيث تعتبر الفقرات التي تم حذفها بهذه الطريقة دالــة إحصــائيًا بالاعتماد على مؤشرات CDIF & ومؤشراتDTF. ولمزيد من التوضيح تم رســم منحنيــات خصائص الفقرة للمجموعتين (المرجعية والمستهدفة) علــى اختبــاري العلــوم والرياضــيات بالاستعانة ببرنامج (SPSS)، للاستدلال على الأداء التفاضلي للفقرة، وبيان الاخــتلاف بــين منحنيات خصائص الفقرة على الاختبارين عبر المجموعتين على مستوى الفقرة الواحدة، وبيان نوع الأداء التفاضلي فيما إذا كان منتظماً أو غير منتظم.

الفصل الرابع

النتائج

هدفت هذه الدراسة إلى الكشف عن الأداء التفاضلي في اختباري العلوم والرياضيات مـــن الاختبار الوطني لضبط نوعية التعليم للصف العاشر وفقراتهما تبعًا لطريقـــة تقـــديم الاختبـــار (ورقي، محوسب)، ويتناول هذا الفصل عرضًا للنتائج التي توصلت إليها الدراسة مبوبة حســـب أسئلتها.

بالرغم من أن فقرات الاختبار الوطني لضبط نوعية التعليم في مجالي العلوم والرياضيات 2011/2010 تم إعدادها من قبل فريق وطني متخصص، تضمن لجان من مشرفين تربويين من إدارة الامتحانات والاختبارات، وأعضاء من إدارة المناهج والكتب المدرسية، ومشرفين تربويين من إدارة التدريب والتأهيل والإشراف التربوي، وعددًا من مشرفي الميدان، إلا أنه تم التأكد من الافتراضات الأساسية الواجب توفرها في البيانات، والمتمثلة بأُحادية البعد، باستخدام التحليل العاملي للمكونات الرئيسية، والاستقلال الموضعي، بحساب مؤشر Z_{Q3}، كما تم التحقق من المطابقة الإحصائية الخاصة بالأفراد والفقرات مع النموذج اللوجستي الثلاثي المعلمة باستخدام برنامج (BILOG–MG 3 تمهيداً لاستخدام طريقة (Dfit) للكشف عن الأداء التفاضلي، وبناءً على ذلك تم الاعتماد على الأفراد والفقرات المطابقين كأساس للكشف عن الفقرات التي تظهر فيها دلالات الأداء التفاضلي تبعا لشكل تقديم الاختبار.

النتائج المتعلقة بسؤال الدراسة الأول:

"هل تظهر فقرات اختبار الرياضيات الوطني لضبط نوعية التعليم لطلبة الصف العاشر(2011/2010) أداءً تفاضليًا ذا دلالة إحصائية عند مستوى الدلالة (α=0.05) يعود لشكل تقديم الاختبار (ورقي مرجعي، محوسب مستهدف)؟".

وللإجابة عن هذا السؤال فقد تم تقسيم بيانات الدراسة إلى بيانات المجموعة المرجعية الخاصة بالشكل الورقي للاختبار، وبيانات المجموعة المستهدفة المتعلقة بالشكل المحوسب للاختبار، ومن ثم تقدير معالم الفقرة (الصعوبة b والتمييز a والتخمين c)، من خلال إجابات المفحوصين عن فقرات الاختبار لكل مجموعة على حدة، باستخدام طريقة الأرجحية العظمى الهامشية (**M**arginal **M**aximum **L**ikelihood, MML) وفقًا للنموذج الثلاثي المعلمة في نظرية الاستجابة للفقرة، عن طريق برنامج (BILOG–MG 3) كما يتضح من الجدولين التاليين:

الجدول (13)
معالم الفقرات المقدرة للمجموعة المرجعية (الشكل الورقي)
بطريقة الأرجحية العظمى الهامشية

c	a	b	رقم الفقرة
0.2534	1.6000	0.3602	1
0.1469	1.8847	0.6305	2
0.2337	3.7430	0.8959	3
0.2756	2.9100	1.5735	4
0.1949	2.6354	0.9442	5
0.2681	1.3102	0.6192	6
0.2429	1.6307	0.5489	8
0.153	2.7894	1.2792	9
0.2179	3.0598	0.6773	10
0.1941	1.9609	1.3103	12
0.4712	1.7027	1.1898	13
0.2799	2.3202	1.4055	14
0.3092	1.9052	1.6490	15
0.1785	2.6275	1.0099	16
0.1889	4.3123	1.2158	19
0.1684	1.7085	0.5129	20
0.2629	3.4676	1.1814	21
0.1777	2.2661	1.2122	23
0.2176	2.0009	1.1680	24
0.2398	1.8320	1.0324	25
0.3460	2.5492	1.4517	26
0.2065	3.5017	1.2400	27
0.3410	1.3142	0.9143	28
0.1792	1.4016	0.6318	29

الجدول (14)

معالم الفقرات وتبايناتها، وتبايناتها المشتركة المقدرة للمجموعة المستهدفة (الشكل المحوسب)

C(a, c)	C(b, c)	C(b, a)	V(c)	V(a)	V(b)	c	a	b	رقم الفقرة
0.0081	0.0018	0.0173	0.0018	0.0874	0.0156	0.2729	1.7888	0.6047	1
0.0051	0.0005	0.0083	0.0011	0.0693	0.0075	0.1448	2.0310	0.5167	2
0.0057	-0.0038	-0.0011	0.0006	0.2730	0.0104	0.2813	2.5173	1.3942	3
0.0056	-0.0039	-0.0058	0.0006	0.2321	0.0114	0.2376	2.3269	1.4747	4
0.0059	-0.0022	-0.0057	0.0008	0.1168	0.0161	0.1963	1.7231	1.4452	5
0.0068	0.0048	0.0129	0.0024	0.0420	0.0236	0.1808	1.2902	0.6673	6
0.0082	0.0006	0.0155	0.0016	0.0999	0.0160	0.2755	1.7597	0.8203	8
0.0038	-0.0021	-0.0071	0.0005	0.1230	0.0099	0.1385	2.0552	1.4058	9
0.0048	-0.0014	0.0066	0.0006	0.1472	0.0050	0.1694	2.6713	0.7847	10
0.0053	-0.0035	-0.0091	0.0006	0.1953	0.0130	0.2150	2.1464	1.5268	12
0.0098	0.0100	0.0261	0.0040	0.0487	0.0376	0.2592	1.3100	0.3192	13
0.0064	-0.0023	0.0044	0.0009	0.1510	0.0123	0.2665	2.0238	1.1929	14
0.0067	-0.0014	-0.0162	0.0011	0.1077	0.0349	0.2403	1.3980	1.7210	15
0.0052	-0.0030	0.0052	0.0005	0.2613	0.0058	0.2341	2.9516	1.0879	16
0.0042	-0.0029	0.0013	0.0004	0.2824	0.0053	0.2006	3.0979	1.2122	19
0.0049	0.0004	0.0063	0.0010	0.0602	0.0091	0.1353	1.7818	0.7272	20
0.0047	-0.0045	0.0030	0.0004	0.6023	0.0059	0.2487	3.7000	1.4272	21
0.0042	-0.0029	-0.0014	0.0004	0.2229	0.0064	0.1775	2.7341	1.2852	23
0.0049	-0.0035	0.0025	0.0005	0.3224	0.0059	0.2269	3.1258	1.2472	24
0.0075	-0.0008	0.0013	0.0013	0.1029	0.0225	0.2534	1.5111	1.3514	25
0.0064	-0.0040	-0.0089	0.0007	0.2145	0.0163	0.2667	2.0635	1.5514	26
0.0045	-0.0034	-0.0011	0.0004	0.2869	0.0067	0.2083	2.8997	1.3451	27
0.0080	-0.0013	-0.0148	0.0012	0.1337	0.0439	0.3276	1.4021	1.7234	28
0.0075	-0.0021	0.0092	0.0010	0.1635	0.0141	0.3192	2.0083	1.1271	29

يلاحظ من الجدولين (13 ,14) أن قيم معالم الفقرات تتباين من حيث الصعوبة والتمييز والتخمين عبر المجموعتين (المستهدفة والمرجعية)، حيث تتراوح قيم معلمة الصعوبة بين (0.3602 إلى 1.6490) للمجموعة المرجعية (الشكل الورقي)، وبين (0.3192 إلى 1.7234) للمجموعة المستهدفة (الشكل المحوسب)، أما معلمة التمييز فتتراوح قيمها بين (1.3102 إلى 4.3123) للمجموعة المرجعية، وبين (1.2902 إلى 3.7000) للمجموعة المستهدفة، في حين أن قيمة التخمين تراوحت بين (0.1469 إلى 0.4712) للمجموعة المرجعية، و(0.1353 إلى 0.3276) للمجموعة المستهدفة، وهذا لا يتفق مع ما أشار إليه هامبلتون وسواميناثان (Hambleton & Swaminathan, 1989) بأنه عندما تتراوح قيم معلمة الصعوبة بين (−2 إلى +2) وقيم معلمة التمييز بين (0.4 إلى 2)، وقيم التخمين بين (صفر إلى 0.25) فإن المعالم تكون مقبولة ومتسقة، وعند مقارنة نتائج النموذج الثلاثي المعلمة

المستخدمة في الدراسة الحالية مع هذه القيم يتبين بأن الفقرات اغلبها لا تتفق مع هذه القيم خاصة فيما يتعلق بقيم: معلمة التمييز، ومعلمة التخمين، حيث أن غالبية هذه القيم تكون أعلى من المدى المحدد في المرجع أعلاه.

في ضوء هذه النتائج، وبهدف تحقيق متطلبات الإجابة عن سؤال الدراسة، تم استخدام برنامج (Equate v2.1) (Baker, 1993)؛ لوضع جميع تقديرات المعالم على مقياس مشترك (Common Metric)، بوضع معالم الفقرات ومعالم القدرة المقدرة من المجموعة المستهدفة على مقياس المجموعة المرجعية بحساب الثابت الضربي (Multiplicative Constant) والجمعي (Additive Constant)، من خلال عملية المعادلة؛ للتقليل من مجموع مربعات الفرق ما بين معالم الفقرة المتقابلة عبر المجموعتين كما يتضح من الجدول (15).

الجدول (15)

معالم الفقرات المقدرة للمجموعة المستهدفة بعد إجراء عملية المعادلة

c	a	b	رقم الفقرة
0.2534	1.5762	0.5041	1
0.1469	1.8567	0.7785	2
0.2337	3.6873	1.0478	3
0.2756	2.8667	1.7356	4
0.1949	2.5962	1.0969	5
0.2681	1.2907	0.7669	6
0.2429	1.6064	0.6956	8
0.1530	2.7479	1.4369	9
0.2179	3.0143	0.8259	10
0.1941	1.9318	1.4685	12
0.4712	1.6773	1.3462	13
0.2799	2.2857	1.5651	14
0.3092	1.8768	1.8123	15
0.1785	2.5885	1.1636	16
0.1889	4.2482	1.3725	19
0.1684	1.6831	0.6591	20
0.2629	3.4160	1.3377	21
0.1777	2.2324	1.3689	23
0.2176	1.9711	1.3241	24
0.2398	1.8048	1.1864	25
0.3460	2.5113	1.6120	26
0.2065	3.4496	1.3971	27
0.3410	1.2946	1.0665	28
0.1792	1.3807	0.7797	29

يتبين من الجدول (15) أن معالم الفقرات المقدرة من المجموعة المستهدفة قد تم وضعها على مقياس المجموعة المرجعية، من خلال عملية التحويل الخطي، حيث تراوحت قيم معلمة الصعوبة بين (0.5041 إلى 1.8123)، أما معلمة التمييز فقد تراوحت بين (1.2907 إلى 4.2482) في حين أن قيم التخمين تراوحت بين(0.1469 إلى 0.4712).

وكون عملية المعادلة تتطلب وجود مجموعة من الفقرات المشتركة (Common Anchor Items) خالية من الأداء التفاضلي لحساب معاملات المعادلة، وبما أن هذا لن يكون ممكنًا قبل الكشف عن الأداء التفاضلي؛ لذلك فإنه تم التغلب على هذه المشكلة، بإجراء المعادلة على مرحلتين 2-Stage Linking (Candell & Drasgow, 1988)، حيث تهدف المرحلة الأولى إلى حساب قيم الثابت الضربي (ميل خط الانحدار)، والثابت الجمعي (ثابت الانحدار) وتنقيتها من الأداء التفاضلي، وذلك بإدخال جميع فقرات اختبار الرياضيات البالغ عددها (24) فقرة كفقرات مشتركة بين المجموعتين في عملية المعادلة الأولية، ويبين الجدول (16) قيم ميل خط الانحدار وثابت الانحدار التي تم الحصول عليها من عملية المعادلة الأولية:

الجدول (16)

قيم الميل وثابت الانحدار من عملية المعادلة الأولية

باستخدام برنامج (Equate v2.1)

الثابت الضربي (الميل)	**1.0151**
الثابت الجمعي (ثابت الانحدار)	**0.1384**

يلاحظ من الجدول (16) أن قيمة الثابت الضربي (ميل خط الانحدار) بلغت (1.0151)، بينما بلغت قيمة الثابت الجمعي (ثابت الانحدار) (0.1384).

بعد ذلك تم إدخال البيانات إلى برنامج (Dfit v8.04)؛ لتحديد الفقرات التي تبدي أداء تفاضليًا كبيرًا؛ لاستثنائها من فقرات الجذع المشترك وليس من ملف معالم الفقرات؛ بهدف تحسين قيم ميل خط الانحدار وثابت الانحدار. حيث قام البرنامج باستخراج قيم "مؤشر الأداء التفاضلي غير التعويضي للفقرة (NCDIF)؛ للكشف عن الأداء التفاضلي في فقرات الاختبار الذي يحسب عن طريق المعادلة رقم (9) المشار لها في الفصل الأول، وذلك كما هو مبين في الجدول (17).

الجدول (17)

قيم مؤشرات NCDIF، CDIF والتباين المشترك C(d, D)عند مستوى الدلالة المحدد

Sig.	NCDIF	CDIF	C(d, D)	SD (d)	Mean (d)	رقم الفقرة
ns	0.00031	0.00035	-0.001	0.014	-0.01	**1**
ns	0.00437	0.00147	0.006	0.041	0.052	**2**
ns	0.00599	-0.00724	-0.007	0.077	0.001	**3**
ns	0.00224	0.00522	0.006	0.047	0.005	**4**
ns	0.0044	-0.00267	-0.005	0.061	-0.026	**5**
ns	0.00238	0.00688	0.003	0.027	-0.041	**6**
ns	0.00032	-0.00181	-0.002	0.017	-0.004	**8**
ns	0.00082	0.00232	0.003	0.027	0.011	**9**
ns	0.00082	0.00427	0.002	0.019	-0.022	**10**
ns	0.0002	-0.00227	-0.002	0.014	0.003	**12**
0.005	**0.01093**	0.01395	0.014	0.104	0.005	**13**
ns	0.00463	0.00331	0.007	0.051	0.045	**14**
ns	0.00144	0.00615	0.004	0.026	-0.028	**15**
ns	0.00276	-0.00415	0	0.01	0.052	**16**
ns	0.00415	0.00219	0.006	0.045	0.046	**19**
ns	0.00174	0.00376	0	0.015	-0.039	**20**
ns	0.00104	0.00002	-0.002	0.018	-0.027	**21**
ns	0.00048	0.00027	0	0.022	0.001	**23**
ns	0.00157	-0.00088	-0.001	0.039	-0.006	**24**
ns	0.0007	-0.00234	-0.002	0.026	0.001	**25**
ns	0.0033	0.00759	0.003	0.024	-0.052	**26**
ns	0.00059	0.00107	0.002	0.018	0.016	**27**
ns	0.01009	0.00141	-0.006	0.045	-0.09	**28**
ns	0.00376	-0.00953	-0.008	0.057	0.022	**29**

يتبين من الجدول (17) أن الفقرة رقم 13 أظهرت أداءً تفاضلياً دال إحصائياً، حيث بلغت قيمة مؤشرNCDIF (0.01093)، وهي قيمة أعلى من القيمة الحرجة Cut Score عند مستوى

الدلالة (α=0.005)، لدى قيام البرنامج بفحص الدلالة الإحصائية بحساب القيم الحرجة للأداء التفاضلي الخاصة بالاختبار، وكل فقرة من فقراته بالإضافة إلى الوسط الحسابي والانحراف المعياري للقيم الحرجة ضمن طريقة معلمة الفقرة التكرارية ,Item Parameter Replication (IPR) عند مستويات دلالة مختلفة تتراوح بين (0.001 إلى 0.5)، وذلك كما هو مبين في الجدول (18).

الجدول (18)

قيم درجات القطع عند مختلف مستويات الدلالة α للفقرات والاختبار

SD	**Mean**	**0.5**	**0.1**	**0.05**	**0.01**	**0.005**	**0.001**	**رقم الفقرة**
0.00234	0.00237	0.0017	0.00529	0.00701	0.01135	0.0128	0.01978	**1**
0.00217	0.00203	0.00127	0.00473	0.00636	0.00967	0.01139	0.01565	**2**
0.00266	0.00244	0.00159	0.00548	0.00737	0.01223	0.01584	0.01692	**3**
0.0026	0.00257	0.00174	0.00609	0.00781	0.01192	0.01375	0.01948	**4**
0.00371	0.00303	0.00171	0.00748	0.01074	0.01694	0.02106	0.02858	**5**
0.00169	0.0018	0.00138	0.00357	0.00471	0.00839	0.00966	0.0133	**6**
0.00278	0.00266	0.00171	0.00618	0.00816	0.01313	0.01459	0.01776	**8**
0.00245	0.00219	0.00139	0.0048	0.00706	0.01162	0.01248	0.02043	**9**
0.00243	0.0023	0.00155	0.00521	0.00749	0.01097	0.01387	0.01555	**10**
0.00295	0.00258	0.0017	0.00549	0.00782	0.01329	0.0161	0.0233	**12**
0.00174	0.00191	**0.00142**	**0.00393**	**0.00551**	**0.00784**	**0.00943**	0.01392	**13**
0.0037	0.00297	0.00173	0.00704	0.01005	0.01671	0.0207	0.03213	**14**
0.00301	0.0028	0.00184	0.00603	0.0083	0.01548	0.01676	0.0235	**15**
0.00236	0.00227	0.00152	0.00527	0.00684	0.01134	0.01296	0.01717	**16**
0.00201	0.00195	0.00132	0.00418	0.00558	0.01017	0.01197	0.0166	**19**
0.00221	0.00211	0.00136	0.005	0.00661	0.01006	0.01084	0.01727	**20**
0.00205	0.00203	0.00142	0.00433	0.00605	0.01027	0.01158	0.01379	**21**
0.0022	0.00208	0.00132	0.00469	0.00653	0.01098	0.01237	0.01642	**23**
0.00225	0.00213	0.00148	0.00459	0.0061	0.0117	0.01235	0.01917	**24**
0.00349	0.00297	0.00183	0.00683	0.00941	0.01764	0.02145	0.02593	**25**
0.00405	0.00312	0.00194	0.00676	0.00983	0.01917	0.02661	0.03192	**26**
0.00203	0.00204	0.00138	0.00452	0.00616	0.00937	0.01154	0.01505	**27**
0.0034	0.00312	0.00198	0.00684	0.01019	0.01601	0.01786	0.01971	**28**
0.00431	0.00347	0.00189	0.00856	0.01145	0.021	0.02314	0.03106	**29**
0.06011	0.05979	0.04079	0.13502	0.17469	0.28662	0.3059	0.45161	DTF

في ضوء ذلك تصبح متغيرات المعادلة (الثابت الضربي والثابت الجمعي) مشككاً فيها وتتلوث بوجود الفقرة رقم 13 نتيجة إظهارها للأداء التفاضلي، كون وجود الأداء التفاضلي يؤثر على دقة تقدير المعالم ودقة المعادلة، وقد اتضح ذلك من خلال قيمة معامل الارتباط بين العلامة الحقيقية للمجموعة المرجعية والمجموعة المستهدفة، والذي يدل على مدى دقة المعادلة، حيث

بلغت قيمة معامل الارتباط (0.99962)؛ لذلك تم استثناء هذه الفقرة من فقرات الجذع المشترك التي دخلت المرحلة الثانية من المعادلة، وليس من جسم الاختبار (ملف معالم الفقرات)؛ للعمل على تحسين قيم الثابت الضربي والثابت الجمعي نتيجة عدم دخول هذه الفقرة في حساب قيم هذه المتغيرات .

كما قام البرنامج بحساب الأداء التفاضلي لاختبار الرياضيات من خلال استخراج قيمة "مؤشر الأداء التفاضلي للاختبار (DTF) " كما يظهر بالجدول (19).

الجدول (19)

معامل الارتباط بين العلامة الحقيقية للمجموعة المستهدفة والمجموعة المرجعية والأداء التفاضلي للاختبار

0.99962	Correlation between focal & reference true scores
0.02934	Differential test functioning (DTF)

يلاحظ من الجدول أن قيمة الأداء التفاضلي للاختبار (0.02934) وهي قيمة تعبر عن المجموع التراكمي لقيم "مؤشر الأداء التفاضلي التعويضي (CDIF)" لجميع الفقرات الداخلة في الاختبار بناء على المعادلة (14) المشار لها في الفصل الأول من الدراسة، وهذه قيمة أقل من القيمة الحرجة عند جميع مستويات الدلالة الإحصائية، وبالتالي لا يبدي الاختبار أداءً تفاضليًا.

تم إدخال البيانات مرة أخرى إلى برنامج (Equate v2.1)، لإجراء المرحلة الثانية من المعادلة، وذلك بتكرار الخطوات السابقة بوضع جميع تقديرات معالم فقرات الجذع المشترك المتبقية التي لم تظهر أداءً تفاضليًا من المرحلة الأولى على مقياس مشترك (Common Metric)، واستثناء الفقرة رقم (13) من فقرات الجذع المشترك، كما يتضح من الجدول (20)؛ لإعادة حساب معاملات المعادلة (الثابت الضربي والثابت الجمعي)؛ لاستخدامها لحساب مؤشرات (DFIT) النهائية. حيث بلغت قيمة الثابت الضربي (**1.0151**) وهي قيمة مشابهة لقيمة

الثابت الضربي التي تم الحصول عليها من عملية المعادلة الأولية، بينما بلغت قيمة الثابت الجمعي (**0.1072**) وهي قيمة أقل من قيمة الثابت الجمعي التي تم الحصول عليها من عملية المعادلة الأولية، وبالتالي تم تنقية قيم هذه المتغيرات من الأداء التفاضلي، وأصبحت هذه المعاملات أكثر استقرارًا، وقد اتضح ذلك من خلال قيمة معامل الارتباط بين العلامة الحقيقية للمجموعة المرجعية والمجموعة المستهدفة التي تم حسابها من خلال برنامج (Dfit v8.04)، والتي تدل على مدى دقة المعادلة، حيث بلغت قيمة معامل الارتباط (0.9999)، وبالتالي هناك ارتفاع قليل على قيمة معامل الارتباط خاصة في الخانة العشرية الرابعة بمقدار (0.00028)، مما يدل على مدى التحسن بدقة المعادلة نتيجة استثناء الفقرة المشتركة التي أبدت أداءً تفاضلياً في المرحلة الأولية من المعادلة، ولم تدخل كفقرة مشتركة في المرحلة الثانية من المعادلة.

الجدول (20)

معالم الفقرات المقدرة للمجموعة المستهدفة بعد إجراء المرحلة الثانية من عملية المعادلة

c	a	b	رقم الفقرة
0.2534	1.5762	0.4729	1
0.1469	1.8567	0.7473	2
0.2337	3.6873	1.0166	3
0.2756	2.8667	1.7044	4
0.1949	2.5962	1.0657	5
0.2681	1.2907	0.7357	6
0.2429	1.6064	0.6644	8
0.1530	2.7479	1.4057	9
0.2179	3.0143	0.7947	10
0.1941	1.9318	1.4373	12
0.4712	1.6773	1.3150	13
0.2799	2.2857	1.5339	14
0.3092	1.8768	1.7811	15
0.1785	2.5885	1.1324	16
0.1889	4.2482	1.3413	19
0.1684	1.6831	0.6279	20
0.2629	3.4160	1.3065	21
0.1777	2.2324	1.3377	23
0.2176	1.9711	1.2929	24
0.2398	1.8048	1.1552	25
0.3460	2.5113	1.5808	26
0.2065	3.4496	1.3659	27
0.341	1.2946	1.0353	28
0.1792	1.3807	0.7485	29

يلاحظ من الجدول (20)، أن معالم الفقرات (التمييز a والصعوبة b والتخمين c) المقدرة من المجموعة المستهدفة قد تم وضعها على مقياس المجموعة المرجعية، من خلال عملية التحويل الخطي، حيث تراوحت قيم معلمة الصعوبة بين (0.4729 إلى 1.7811)، أما معلمة التمييز فقد تراوحت بين (1.2907 إلى 4.2482)، في حين أن قيم التخمين تراوحت بين (0.1469 إلى 0.4712). وفي ضوء ذلك يلاحظ

أن قيم معلمة التمييز والتخمين لم تتغير عما كانت عليه في مرحلة المعادلة الأولية في حين حدث تناقص على قيم معلمة الصعوبة؛ بسبب استثناء الفقرة13 التي أبدت أداءً تفاضليًا من الفقرات المشتركة التي دخلت المرحلة الثانية من المعادلة، مما زاد من دقة مُعادلة هذه المعلمة.

وبإعادة استخدام برنامج (Dfit v8.04) للمرة الثانية، تم استخراج قيم مؤشر "الأداء التفاضلي غير التعويضي للفقرة (NCDIF)" للكشف عن الأداء التفاضلي في فقرات الاختبار، حيث تظهر الفقرة أداءً تفاضلياً لمتغير طريقة تقديم الاختبار إذا كانت قيمة NCDIF المحسوبة أعلى من القيمة الحرجة، لدى قيام البرنامج بفحص الدلالة الإحصائية بحساب القيم الحرجة للأداء التفاضلي الخاصة بالاختبار وكل فقرة من فقراته، بالإضافة إلى الوسط الحسابي والانحراف المعياري للقيم الحرجة الخاصة بها ضمن طريقة معلمة الفقرة التكرارية Item (Parameter Replication, IPR) عند مستويات دلالة مختلفة تتراوح بين (0.001 إلى 0.5)، كما يتضح بالجدول (21). (القيمة الحرجة للاختبار عبارة عن مجموع القيم الحرجة لـــ NCDIF لجميع فقرات الاختبار).

الجدول (21) قيم درجات القطع عند مختلف مستويات الدلالة α للفقرات والاختبار

SD	Mean	0.5	0.1	0.05	0.01	0.005	0.001	رقم الفقرة
0.00245	0.00239	0.00164	0.00554	0.00727	0.01146	0.01437	0.01581	1
0.00212	0.00202	0.00136	0.00449	0.00599	0.01076	0.01253	0.01449	2
0.00292	0.00258	0.00157	0.00599	0.00847	0.01398	0.01465	0.0223	3
0.00268	0.00253	0.00177	0.00576	0.00743	0.01204	0.01428	0.02184	4
0.00357	0.00307	0.00172	0.00795	0.01069	0.01601	0.01734	0.02672	5
0.00181	0.00195	0.00138	0.00444	0.00564	0.00792	0.00961	0.01206	6
0.00315	0.00285	0.00177	0.00639	0.00892	0.01671	0.01798	0.01886	8
0.0022	0.00225	0.0016	0.00477	0.00638	0.01091	0.01241	0.01421	9
0.00226	0.00218	0.00145	0.00501	0.00652	0.01145	0.01329	0.01597	10
0.00288	0.00269	0.0018	0.006	0.00779	0.01341	0.01622	0.02324	12
0.00182	0.00189	**0.00134**	**0.00409**	**0.00515**	**0.00753**	**0.00953**	0.01524	13
0.00361	0.00285	0.00161	0.00677	0.00942	0.01553	0.02041	0.03023	14
0.00332	0.00284	0.00178	0.00637	0.00851	0.01719	0.02106	0.02669	15
0.00206	0.00212	0.0015	0.00476	0.00636	0.00907	0.01009	0.01492	16
0.00198	0.00202	0.00135	0.00471	0.0062	0.00862	0.01052	0.01284	19
0.00264	0.00226	0.00144	0.00474	0.00716	0.01315	0.01744	0.02098	20
0.00214	0.00201	0.00139	0.00418	0.00617	0.01041	0.01216	0.01468	21
0.00229	0.00213	0.00138	0.00497	0.00633	0.0103	0.01296	0.01705	23
0.0022	0.00217	0.00149	0.00477	0.00646	0.01013	0.01174	0.01649	24
0.00302	0.00282	0.00182	0.00662	0.00862	0.01445	0.01673	0.02033	25
0.00417	0.0033	0.00191	0.00759	0.01104	0.02	0.02579	0.03387	26
0.00227	0.00213	0.00147	0.00466	0.00614	0.00987	0.01291	0.01733	27
0.00391	0.00321	**0.00188**	**0.00753**	**0.01017**	0.01695	0.01978	0.04509	28
0.00389	0.00317	0.0018	0.00758	0.01055	0.01785	0.02341	0.0296	29
0.05766	0.05987	**0.0422**	0.13476	0.17205	0.28782	0.35639	0.39239	DTF

الجدول (22)

النتائج النهائية للأداء التفاضلي لفقرات اختبار الرياضيات عبر المجموعتين المرجعية والمستهدفة من خلال قيم مؤشراتNCDIF، CDIF، والتباين المشترك C(d, D)عند مستوى الدلالة المحدد

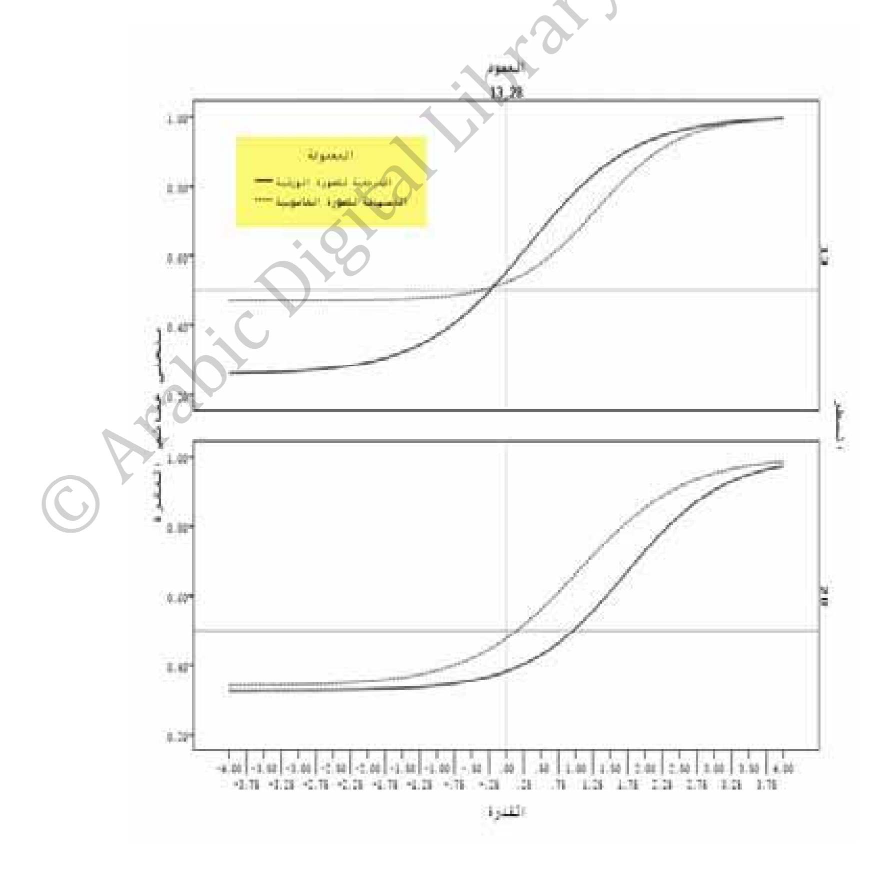
Sig.	NCDIF	CDIF	C(d, D)	SD (d)	Mean (d)	رقم الفقرة
ns	0.00054	0.00253	-0.001	0.017	-0.016	1
ns	0.00341	-0.00681	0.002	0.037	0.045	2
ns	0.00685	-0.00047	-0.001	0.083	-0.004	3
ns	0.00185	0.00097	0.001	0.043	0.002	4
ns	0.00522	0.00561	-0.001	0.065	-0.032	5
ns	0.00271	0.01008	0.001	0.025	-0.046	6
ns	0.00051	0.0007	-0.001	0.021	-0.01	8
ns	0.00067	0.0003	0.002	0.025	0.006	9
ns	0.00099	0.00632	0.001	0.014	-0.028	10
ns	0.00031	-0.00060	-0.001	0.018	-0.001	12
0.005	**0.01046**	0.00509	0.005	0.102	0.002	13
ns	0.00396	-0.00602	0.002	0.048	0.041	14
ns	0.00149	0.00735	0.001	0.024	-0.03	15
ns	0.00221	-0.00950	0	0.009	0.046	16
ns	0.00333	-0.00589	0.002	0.041	0.041	19
ns	0.00231	0.00864	0	0.016	-0.045	20
ns	0.00153	0.00532	-0.001	0.024	-0.031	21
ns	0.0004	0.00005	-0.001	0.02	-0.004	23
ns	0.00162	0.00022	-0.002	0.039	-0.01	24
ns	0.00086	0.00023	-0.001	0.029	-0.004	25
ns	0.00347	0.01193	0.001	0.02	-0.055	26
ns	0.00037	-0.00139	0.001	0.015	0.012	27
0.05	**0.01101**	0.01636	-0.002	0.047	-0.094	28
ns	0.00387	-0.00684	-0.004	0.06	0.017	29

وفقا لمؤشر NCDIF يتبين من الجدول (22) أن فقرتين من فقرات اختبار الرياضيات قد أظهرت أداءً تفاضلياً تبعًا لمتغير شكل تقديم الاختبار (ورقي، محوسب) بنسبة 8.33% تقريباً من أصل 24 فقرة تكون منها الاختبار بصورته النهائية، منها الفقرة 13 التي استمرت بإبداء الأداء التفاضلي في المرحلة الثانية من المعادلة عند مستوى الدلالة (α=0.005)، حيث بلغت قيمة مؤشر (NCDIF) (0.01046)، وهي أقل من قيمة مؤشر NCDIF التي حسبت في المرحلة الأولية من المعادلة، علمًا بأن هذه الفقرة تتعلق بمحور الجبر، وتهدف إلى استخدام مفهوم المصفوفة المنفردة في إيجاد مجاهيل، حيث تتطلب إيجاد رتبة المصفوفة المبينة في السؤال، ومن جهة أخرى أظهرت الفقرة28 أداءً تفاضليًا عند مستوى الدلالة (α=0.05)، حيث بلغت قيمة مؤشر NCDIF (0.01101)، وهي فقرة تتعلق بمحور الإحصاء والاحتمالات، وتهدف إلى استخدام قوانين الاحتمالات في حساب احتمالات حوادث معينة، حيث تتطلب إيجاد احتمال الحصول على عددين مجموعهما يقبل القسمة على (6) في تجربة إلقاء حجري نرد منتظمين مرة واحدة، علمًا بأن قيمة مؤشر NCDIF لهاتين الفقرتين أعلى من القيمة الحرجة عند مستوى الدلالة ألفا (α) المحدد أعلاه عند فحص الدلالة الإحصائية بطريقة (IPR).

كما تبين أن (22) فقرة من أصل 24 فقرة تكون منها الاختبار، لم تظهر أداءً تفاضليًا لمتغير شكل تقديم الاختبار بنسبة 91.67% تقريباً، حيث كانت قيمة مؤشرNCDIF أقل من القيمة الحرجة عند جميع مستويات الدلالة ألفا (α)، لدى فحص الدلالة الإحصائية بطريقة (IPR)، حيث أن قيمة مؤشر NCDIF للمجموعتين تكاد لا تؤخذ بعين الاعتبار مقارنة بالقيمة الحرجة (درجة القطع) التي تم اعتمادها معياراً للحكم على وجود الأداء أو عدم وجوده، ويتضح

ذلك في الفقرات: (24, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 10, 9, 8, 7, 6, 5, 4, 3, 2,1).

كما تم التأكد من مدى إظهار الفقرتين للأداء التفاضلي من خلال رسم منحنيات خصائص الفقرة للمجموعتين المستهدفة والمرجعية باستخدام البرمجية الإحصائية (SPSS) للنموذج الثلاثي المعلمة؛ للاستدلال على الأداء التفاضلي للفقرة؛ ولبيان الاختلاف بين منحنيات خصائص الفقرة للمجموعتين (المستهدفة محوسب، والمرجعية ورقي) على اختبار الرياضيات على مستوى الفقرة الواحدة، وبيان نوع الأداء التفاضلي فيما إذا كان منتظماً أو غير منتظماً كما يتضح في الشكل (4).



شكل (4): منحنى خصائص الفقرة للفقرتين رقم 13 و28 اللتين أظهرتا أداءً تفاضلياً تبعا لشكل تقديم الاختبار

حيث تبين، من خلال تفحص منحنى خصائص الفقرة أن الفقرة (13) تبدي أداءً تفاضـــلياً غير منتظم، حيث أن احتمال الإجابة الصحيحة عن الفقرة يختلف تبعـــاً لمتغيـــر شـــكل تقـــديم الاختبار عند المفحوصين المتساوين في القدرة، وأن منحنى المجموعة المستهدفة أكثر تمييزًا من منحنى المجموعة المرجعية إذا ما تفحصنا بالنظر ميل منحنيي المجموعتين، فاحتمال الإجابـــة الصحيحة عن الفقرة: أعلى للمجموعة المستهدفة (الشكل المحوسب) من المجموعة المرجعيـــة (الشكل الورقي) ضمن مستويات القدرة الدنيا تقريبًا؛ أي أن الفقرة تعمـــل لصـــالح المجموعـــة المستهدفة (المحوسب) عند هذا المستوى من القدرة، واحتمال الإجابة الصحيحة أقل للمجموعـــة المستهدفة من المجموعة المرجعية ضمن مستويات القدرة المرتفعة تقريبًا؛ أي أن الفقرة تعمـــل لصالح المجموعة المرجعية (الورقي) عند هذا المستوى من القدرة.

بينما أظهرت الفقرة (28) أداءً تفاضلياً منتظم، حيث يتضح من الشكل أن احتمال الإجابة الصحيحة عن الفقرة أعلى للمجموعة المستهدفة (المحوسب) من المجموعة المرجعية (الورقي) عند معظم مستويات القدرة، وأن منحنيي خصائص الفقرة للمجموعة المرجعية والمجموعة المستهدفة لهما نفس التمييز تقريبًا (ميل المنحنيين متقارب)، وهذا يعني أن هذه الفقرة تعمل لصالح المجموعة المستهدفة (المحوسب).

وعند تفحص منحنيات خصائص الفقرة للفقرات التي لم تبد أداءً تفاضليًا بالاعتماد على قيم مؤشر NCDIF يلاحظ أن منحنيات خصائص الفقرة لكل فقرة من هذه الفقرات عبر المجموعتين المرجعية والمستهدفة، تعكس مقداراً قليلا من عدم الانتظام في منحنياتها؛ إذا ما قورنت مع الفقرات التي تظهر منحنياتها أداءً تفاضليًا، وقد يكون السبب في ذلك أن أداء المفحوصين عليها من كلا المجموعتين المرجعية والمستهدفة متقاربًا، أو أن الفرق بين المنحنيين

للمجموعتين كان قريبًا من الصفر، بحيث يكاد ينطبق المنحنيان على بعضهما عند جميع مستويات القدرة المختلفة، وبالتالي فالاختلاف بين المنحنيين لا يتعدى معيارًا معينًا وهو خطأ المعاينة.(أنظر الملحق 9)

كما تم استخراج قيم مؤشر الأداء التفاضلي التعويضي للفقرة CDIF، لمعرفة العلاقة بين الأداء التفاضلي للفقرة مع الأداء التفاضلي للاختبار، فعند تفحص قيم مؤشر الأداء التفاضلي التعويضي للفقرة من الجدول (20) تبين أن بعض قيمه سالبة كما في الفقرات (24, 22, 15, 14, 12, 10,3,2)، والبعض الأخر قيمه موجبة كما في الفقرات (23, 21, 20, 19, 18, 17, 16, 13, 11, 9, 8, 7, 6, 5, 4, 1)، وأن أعلى قيمه له (0.01636) للفقرة (28) وأقل قيمة له (-0.0095) للفقرة (16)، وبالتالي تعمل القيم السالبة والموجبة للمؤشر بإلغاء بعضها بعضًا عندما تجمع على مستوى الاختبار، وما يتبقى يعكس مدى إسهام الفقرات في الأداء التفاضلي للاختبار، كما يتضح عند الكشف عن الأداء التفاضلي للاختبار من خلال استخراج قيمة مؤشر الأداء التفاضلي للاختبار DTF وفق الجدول (23)

الجدول (23)

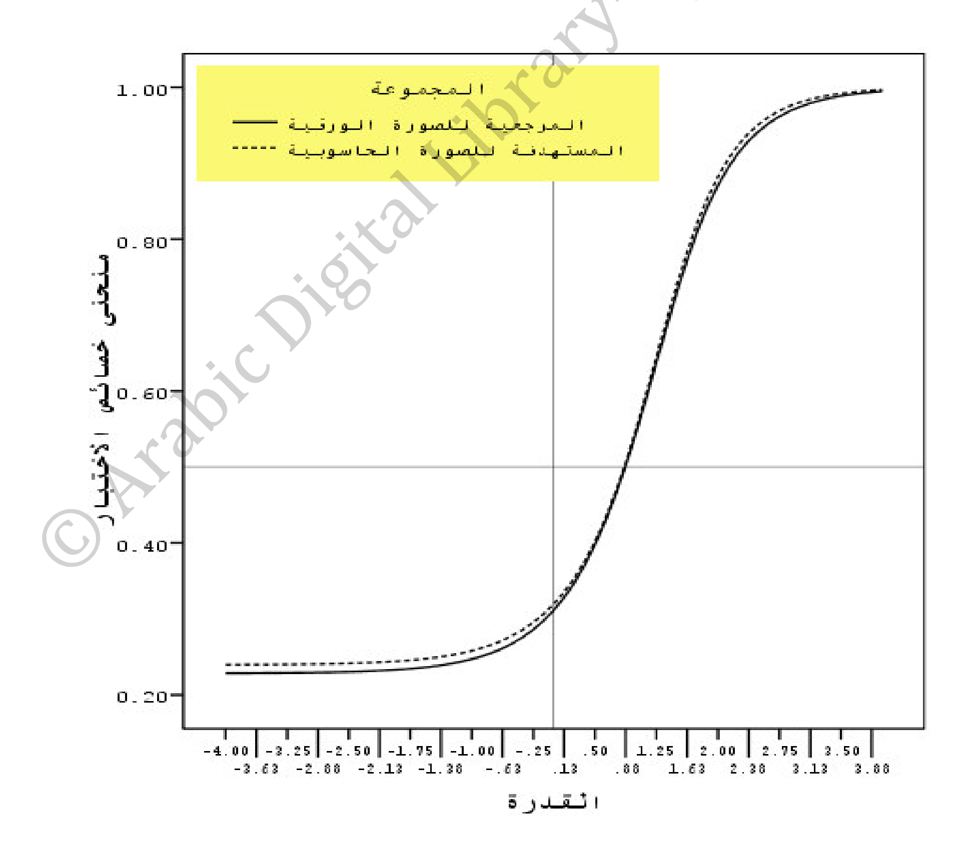
معامل الارتباط بين العلامة الحقيقية للمجموعة المستهدفة والمجموعة المرجعية، والأداء التفاضلي للاختبار

0.9999	Correlation between focal & reference true scores
0.04419	Differential test functioning (DTF)

يتبين من الجدول (23) أن الأداء التفاضلي للاختبار بلغ (0.04419)، وهي قيمة أعلى من القيمة الحرجة عند مستوى الدلالة (α=0.5)، وهذه القيمة تعبر عن المجموع التراكمي لقيم مؤشر CDIF لجميع الفقرات الداخلة في الاختبار، وعند عمل فحص لمؤشر إزالة الفقرة item removal index (2-CDIF–NCDIF) بدءً بالفقرات الأعلى مؤشر إزالة، لتحديد الفقرات التي

ستزال من الاختبار (الفقرات التي تزال توصف بأنها تبدي أداء تفاضلي بالاعتماد على مؤشرات CDIF & ومؤشراتDTF)

تبين عدم حذف أي فقرة من الاختبار، ومرد هذا إلى أنه بالرغم من أن قيمة الأداء التفاضلي للاختبار أعلى من القيمة الحرجة عند مستوى الدلالة (α=0.5) إلا أنه غير دال عمليًا لأنه غير دال إحصائيًا عند مستوى الدلالة (α=0.05)، التي اعتمدت معيارًا للدلالة الإحصائية في هذه الدراسة، وبالتالي الدلالة الإحصائية عند هذه القيمة عديمة القيمة، وقد تم التحقق من ذلك برسم منحنى خصائص الاختبار للمجموعتين المستهدفة والمرجعية باستخدام البرمجية الإحصائية (SPSS) كما يتضح من الشكل (5).



شكل (5): منحنى خصائص الاختبار للمجموعتين المستهدفة والمرجعية

النتائج المتعلقة بسؤال الدراسة الثاني:

"هل تظهر فقرات اختبار العلوم الوطني لضبط نوعية التعليم لطلبة الصف العاشر (2011/2010) أداءً تفاضليًا ذا دلالة إحصائية عند مستوى الدلالة (α=0.05) يعود إلى طريقة تقديم الاختبار (ورقي مرجعي، محوسب مستهدف)؟".

وللإجابة عن هذا السؤال فقد تم تقسيم بيانات الدراسة إلى بيانات المجموعة المرجعية الخاصة بالشكل الورقي للاختبار، وبيانات المجموعة المستهدفة المتعلقة بالشكل المحوسب للاختبار، ومن ثم تقدير معالم الفقرة (الصعوبة b والتمييز a والتخمين c) لكل مجموعة على حدةٍ باستخدام طريقة الأرجحية العظمى الهامشية **M**arginal **M**aximum **L**ikelihood, (MML) وفقًا للنموذج الثلاثي المعلمة في نظرية الاستجابة للفقرة، عن طريق برنامج (BILOG–MG 3) كما يتضح من الجدولين التاليين:

الجدول (24)

معالم الفقرات وتبايناتها، وتبايناتها المشتركة المقدرة للمجموعة المستهدفة (الشكل المحوسب)

رقم الفقرة	b	a	c	V(b)	V(a)	V(c)	C(b,a)	C(b,c)	C(a,c)
1	**-0.0838**	1.1425	0.2536	0.0567	0.0351	0.0059	0.0298	0.0166	0.0094
2	0.3736	0.7957	0.2504	0.1183	0.0245	0.0068	0.0294	0.0267	0.0084
3	0.0804	**0.6526**	0.2541	0.1908	0.016	0.0083	0.0294	0.0372	0.0064
4	**3.0096**	1.7904	**0.134**	0.1763	0.3952	0.0002	-0.1867	-0.0046	0.0034
5	1.5189	1.1091	0.1561	0.0365	0.0491	0.0015	-0.0076	0.0023	0.0057
7	0.5243	0.9955	0.1702	0.0397	0.0255	0.0033	0.0124	0.0095	0.0056
9	2.2002	2.6327	0.2725	0.0322	0.634	0.0004	-0.0347	-0.0064	0.0051
10	1.0239	2.1087	0.2673	0.0108	0.1492	0.0009	0.0043	-0.0018	0.0066
11	1.2285	1.4567	0.3198	0.0288	0.1039	0.0016	0.0055	0.0007	0.0084
12	0.3188	1.8071	0.2337	0.0133	0.0736	0.002	0.0161	0.003	0.0077
13	0.4611	1.0617	0.277	0.0646	0.0448	0.0052	0.0317	0.016	0.0109
14	1.7664	2.5282	0.2069	0.0164	0.322	0.0004	-0.0198	-0.0047	0.0049
15	0.7889	1.2168	0.2526	0.0338	0.0547	0.0029	0.0166	0.0065	0.0086
18	1.7208	**3.0816**	0.2935	0.0151	0.6136	0.0004	-0.0153	-0.0057	0.0055
19	2.0665	2.388	0.2177	0.0271	0.3702	0.0004	-0.0322	-0.0049	0.0046
20	1.2423	1.5689	**0.3207**	0.025	0.1174	0.0014	0.0035	-0.0004	0.008
22	0.8294	2.1486	0.2404	0.0089	0.13	0.001	0.0075	-0.0011	0.0066
23	0.7297	2.5224	0.1493	0.0048	0.1261	0.0007	0.0041	-0.0012	0.0048
24	1.2233	2.2744	0.2879	0.0117	0.2041	0.0007	-0.0009	-0.003	0.0062
25	2.3985	2.698	0.1565	0.032	0.6788	0.0002	-0.0564	-0.0055	0.0038
26	1.0953	1.8873	0.2026	0.0113	0.1076	0.0009	0.0005	-0.0014	0.0057
27	0.9562	2.0237	0.2972	0.0123	0.1434	0.0011	0.0076	-0.0013	0.0072
28	1.6747	1.2441	0.1866	0.0375	0.0723	0.0012	-0.014	0.0002	0.0062
29	2.8339	1.2062	0.1694	0.1829	0.1278	0.0006	-0.1038	-0.0023	0.0047
30	1.8387	1.0085	0.2438	0.0743	0.0675	0.0021	-0.0148	0.0047	0.0082

الجدول (25)

معالم الفقرات المقدرة للمجموعة المرجعية (الشكل الورقي)

بطريقة الأرجحية العظمى الهامشية

c	a	b	رقم الفقرة
0.2976	0.9212	-0.2019	**1**
0.3106	**0.4974**	**-0.2301**	**2**
0.2995	0.6433	0.0906	**3**
0.1180	1.6111	3.0623	**4**
0.0821	1.2608	1.4946	**5**
0.2406	0.8762	0.5435	**7**
0.2208	2.2940	2.1916	**9**
0.2339	2.0114	0.7725	**10**
0.2756	0.9474	1.3280	**11**
0.2305	1.8672	0.1931	**12**
0.4925	2.0716	1.3386	**13**
0.2452	2.2527	1.8550	**14**
0.3370	1.4029	0.8680	**15**
0.2264	1.5945	1.7198	**18**
0.1580	2.0583	2.1690	**19**
0.2975	1.9974	1.1027	**20**
0.2860	2.2793	1.2937	**22**
0.1210	2.7260	0.4676	**23**
0.2946	2.5410	1.4569	**24**
0.1593	**2.9216**	2.1417	**25**
0.1926	1.3151	1.0433	**26**
0.2870	1.8253	0.5855	**27**
0.1931	1.3033	1.3313	**28**
0.2020	1.0748	**3.6807**	**29**
0.1770	1.0350	1.8683	**30**

يلاحظ من الجدول أن قيم معالم الفقرات تتباين من حيث الصعوبة والتمييز والتخمين عبر المجموعتين (المستهدفة محوسب والمرجعية ورقي)، حيث تتراوح قيمة معلمة الصعوبة بين (-0.2301 إلى 3.6807) للمجموعة المرجعية (الشكل الورقي). وبين (-0.0838 إلى 3.0096) للمجموعة المستهدفة (الشكل المحوسب)، أما معلمة التمييز فقد تراوحت بين (0.4974 إلى 2.9216) للمجموعة المرجعية، وبين (0.6526 إلى 3.0816) للمجموعة المستهدفة، في حين أن قيم التخمين تراوحت بين (0.0821 إلى 0.4925) للمجموعة المرجعية، وبين (0.1340 إلى 0.3207) للمجموعة المستهدفة، وهذا لا يتفق مع ما أشار إليه هامبلتون وسواميناثان (Hambleton & Swaminathan, 1989) بأنه عندما تتراوح قيم معلمة

الصعوبة بين (−2 إلى +2)، وقيم معلمة التمييز بين (0.4 إلى 2) وقيم التخمين بين (صفر إلى 0.25) فإن المعالم مقبولة ومتسقة، وعند مقارنة نتائج النموذج الثلاثي المعلمة المستخدمة في الدراسة الحالية مع هذه القيم تبين بأن الفقرات اغلبها لا تتفق مع هذه القيم

في ضوء هذه النتائج وبهدف تحقيق متطلبات الإجابة عن سؤال الدراسة، تم استخدام برنامج (Equate v2.1) (Baker, 1993)، لوضع جميع تقديرات المعالم على مقياس مشترك (Common Metric)، بوضع معالم الفقرات ومعالم القدرة المقدرة من المجموعة المستهدفة على مقياس المجموعة المرجعية بحساب الثابت الضربي (Multiplicative Constant) والجمعي (Additive Constant) من خلال عملية المعادلة؛ للتقليل من مجموع مربعات الفرق بين معالم الفقرة المتقابلة عبر المجموعتين كما يتضح من الجدول(26):

الجدول (26)

معالم الفقرات المقدرة للمجموعة المستهدفة بعد إجراء عملية المعادلة الأولية

c	a	b	رقم الفقرة
0.2976	0.9995	-0.0859	**1**
0.3106	0.5397	**-0.1118**	**2**
0.2995	0.6980	0.1837	**3**
0.1180	1.7480	2.9227	**4**
0.0821	1.3679	1.4778	**5**
0.2406	0.9506	0.6012	**7**
0.2208	2.4888	2.1202	**9**
0.2339	2.1823	0.8122	**10**
0.2756	1.0279	1.3242	**11**
0.2305	2.0258	0.2782	**12**
0.4925	2.2476	1.3340	**13**
0.2452	2.4441	1.8099	**14**
0.3370	1.5220	0.9002	**15**
0.2264	1.7300	1.6854	**18**
0.1580	2.2332	2.0993	**19**
0.2975	2.1671	1.1166	**20**
0.2860	2.4729	1.2926	**22**
0.1210	2.9575	0.5312	**23**
0.2946	2.7568	1.4430	**24**
0.1593	3.1697	2.0742	**25**
0.1926	1.4269	1.0618	**26**
0.2870	1.9803	0.6399	**27**
0.1931	1.4141	1.3272	**28**
0.2020	1.1662	**3.4927**	**29**
0.1770	1.1229	1.8222	**30**

يلاحظ من الجدول(26) أن معالم الفقرات المقدرة من المجموعة المستهدفة قد تم وضعها على مقياس المجموعة المرجعية، من خلال عملية التحويل الخطي، حيث تراوحت قيم معلمة الصعوبة بين (−0.1118 إلى 3.4927)، أما معلمة التمييز فقد تراوحت بين (0.5397 إلى 3.1697)، في حين أن قيم التخمين تراوحت بين (0.0821 إلى 0.4925).

وكون عملية المعادلة تتطلب وجود مجموعة من الفقرات المشتركة (Common Anchor Items) خالية من الأداء التفاضلي، لحساب معاملات المعادلة، وبما أن هذا لن يكون ممكنًا قبل الكشف عن الأداء التفاضلي؛ لذلك فإنه تم التغلب على هذه المشكلة بإجراء المعادلة على مرحلتين 2−Stage Linking (Candell & Drasgow, 1988)، حيث تهدف المرحلة الأولى إلى حساب قيم الثابت الضربي (ميل خط الانحدار) والثابت الجمعي (ثابت الانحدار) وتنقيتها من الأداء التفاضلي، وذلك بإدخال جميع فقرات اختبار العلوم والبالغ عددها 25 فقرة كفقرات مشتركة في عملية المعادلة في المرحلة الأولية، ويبين الجدول (27) قيم الميل وثابت الانحدار التي تم الحصول عليها من عملية المعادلة الأولية:

الجدول (27)

قيم الميل وثابت الانحدار من عملية المعادلة الأولية باستخدام برنامج Equate v2.1

الثابت الضربي (الميل)	**0.9217**
الثابت الجمعي (ثابت الانحدار)	**0.1002**

يلاحظ من الجدول (27) أن قيمة الميل بلغت (0.9217) بينما بلغت قيمة ثابت الانحدار (0.1002)، ولمعرفة مدى تأثر هذه القيم بالأداء التفاضلي للفقرات فقد تم إدخال البيانات إلى برنامج (Dfit v8.04)؛ لتحديد الفقرات التي تبدي أداءً تفاضليًا كبيرًا، لاستثنائها من فقرات الجذع المشترك وليس من ملف معالم الفقرات، بهدف تحسين قيم الميل وثابت الانحدار. حيث قام البرنامج باستخراج قيم "مؤشر الأداء التفاضلي غير التعويضي للفقرة (NCDIF)"للكشف عن

الأداء التفاضلي في فقرات الاختبار الذي يحسب عن طريق المعادلة رقم (9) المشار لها في الفصل الأول من الدراسة، وذلك كما هو مبين في الجدول (28).

الجدول (28)
قيم مؤشرات NCDIF، CDIF والتباين المشترك C(d, D) عند مستوى الدلالة المحدد

Sig.	NCDIF	CDIF	DIF C(d,D)	SD (d)	Mean (d)	رقم الفقرة
ns	0.00101	0.00141	0.002	0.025	-0.019	**1**
0.005	**0.00923**	0.0012	0.002	0.044	-0.085	**2**
ns	0.00018	0.0003	0	0.004	-0.013	**3**
ns	0.00019	0.00008	0	0.007	0.012	**4**
0.05	**0.00698**	0.00123	0.001	0.029	0.078	**5**
ns	0.00165	0.00106	0.001	0.022	-0.034	**7**
ns	0.00206	0.00032	0	0.015	0.043	**9**
ns	0.00153	-0.00203	-0.002	0.037	-0.013	**10**
ns	0.00072	-0.00126	-0.001	0.022	0.015	**11**
ns	0.00027	-0.00062	-0.001	0.015	-0.006	**12**
0.01	**0.00892**	0.00839	0.008	0.094	-0.005	**13**
ns	0.00122	-0.00032	0	0.011	-0.033	**14**
ns	0.00131	0.00197	0.002	0.022	-0.029	**15**
ns	0.00205	-0.0021	-0.002	0.039	0.023	**18**
ns	0.00306	0.00023	0	0.005	0.055	**19**
ns	0.00164	0.00102	0.001	0.034	0.022	**20**
ns	0.00828	0.00475	0.004	0.075	0.051	**22**
ns	0.00364	-0.00289	-0.003	0.054	-0.028	**23**
ns	0.00253	0.00206	0.002	0.037	0.034	**24**
ns	0.00142	0.00054	0.001	0.035	-0.015	**25**
ns	0.0011	-0.00202	-0.002	0.027	-0.02	**26**
ns	0.00431	-0.00343	-0.003	0.041	-0.051	**27**
ns	0.00224	-0.00102	-0.001	0.03	-0.037	**28**
ns	0.00087	0.00022	0	0.028	-0.009	**29**
ns	0.00463	0.00064	0	0.014	0.067	**30**

يلاحظ من الجدول أن الفقرات رقم (2 ,5 ,13) أظهرت أداءً تفاضلياً دال إحصائياً وفقاً لمؤشر NCDIF، حيث بلغت قيم هذا المؤشر (0.00892, 0.00698, 0.00923) على التوالي وهي قيم أعلى من القيمة الحرجة عند مستوى الدلالة (α=0.005, 0.05, 0.01) على التوالي، وذلك لدى قيام البرنامج بفحص الدلالة الإحصائية بحساب القيم الحرجة للأداء التفاضلي الخاصة بالاختبار وكل فقرة من فقراته والوسط الحسابي والانحراف المعياري للقيم الحرجة ضمن طريقة معلمة الفقرة التكرارية (Item Parameter Replication, IPR) عند مستويات دلالة مختلفة تتراوح بين (0.001 إلى 0.5) كما يتضح من الجدول (29).

الجدول (29)

قيم درجات القطع عند مستويات الدلالة α للفقرات والاختبار

SD	Mean	0.5	0.1	0.05	0.01	0.005	0.001	رقم الفقرة
0.00156	0.00173	0.00128	0.00372	0.00489	0.00747	0.0083	0.01006	**1**
0.00152	0.00155	**0.00109**	**0.00339**	**0.00448**	**0.0067**	**0.00833**	0.01082	**2**
0.0015	0.00157	0.0011	0.00336	0.0041	0.0073	0.00872	0.01111	**3**
0.02569	0.00528	0.00128	0.00714	0.01182	0.09656	0.21222	0.34278	**4**
0.00175	0.0019	**0.00138**	**0.00416**	**0.00536**	0.00887	0.01002	0.01132	**5**
0.00138	0.00157	0.00118	0.00335	0.00422	0.00691	0.00726	0.00891	**7**
0.00717	0.0028	0.00137	0.00526	0.0078	0.02154	0.02743	0.10332	**9**
0.00394	0.00313	0.00168	0.0078	0.01096	0.0182	0.02466	0.03057	**10**
0.00305	0.00292	0.0019	0.0066	0.00911	0.01545	0.01606	0.02348	**11**
0.00186	0.00207	**0.00148**	**0.00467**	**0.00578**	**0.0086**	0.00935	0.01087	**12**
0.00186	0.00198	0.00147	0.00426	0.00525	0.00801	0.00919	0.01582	**13**
0.0026	0.00229	0.0015	0.00472	0.00664	0.01299	0.01846	0.02055	**14**
0.00183	0.00217	0.0017	0.00461	0.00577	0.00821	0.00914	0.01244	**15**
0.00274	0.00235	0.00156	0.00473	0.00679	0.01448	0.01817	0.02223	**18**
0.00392	0.00234	0.00148	0.0049	0.0068	0.01147	0.01758	0.06756	**19**
0.00346	0.00314	0.00197	0.00758	0.0106	0.01619	0.01738	0.02026	**20**
0.00331	0.00282	0.0017	0.00681	0.00932	0.01579	0.01776	0.02814	**22**
0.00365	0.00278	0.00144	0.0068	0.00978	0.01583	0.02117	0.02746	**23**
0.00306	0.00285	0.00193	0.00615	0.00899	0.01521	0.01706	0.02237	**24**
0.01456	0.00316	0.00133	0.00453	0.00685	0.03415	0.04659	0.24061	**25**
0.00351	0.00308	0.00183	0.00755	0.01028	0.01609	0.01838	0.02606	**26**
0.00335	0.00297	0.00185	0.00694	0.0093	0.01632	0.01935	0.02228	**27**
0.0026	0.00242	0.00163	0.00558	0.0075	0.01261	0.01519	0.01829	**28**
0.00678	0.003	0.00142	0.00586	0.00914	0.03007	0.04518	0.09605	**29**
0.00228	0.00193	0.00141	0.00396	0.00509	0.01112	0.01466	0.02582	**30**
0.06433	**0.0632**	**0.04439**	**0.13124**	**0.17297**	**0.29712**	**0.37719**	**0.48067**	**DTF**

وبناءً على ذلك فإن متغيرات المعادلة (الثابت الضربي والثابت الجمعي) قد تلوثت وأصبح مشكك فيها بوجود الفقرات رقم (2 ,5 ,13) نتيجة إظهارها للأداء التفاضلي، حيث أن وجود الأداء التفاضلي قد يؤثر على دقة تقدير المعالم ودقة المعادلة؛ وقد اتضح ذلك من خلال قيمة معامل الارتباط بين العلامة الحقيقية للمجموعة المرجعية والمجموعة المستهدفة، والذي يدل على مدى دقة المعادلة حيث بلغت قيمة معامل الارتباط (0.99964)؛ لذلك تم استثناء هذه الفقرات من فقرات الجذع المشترك التي دخلت المرحلة الثانية من المعادلة، وليس من جسم الاختبار (ملف معالم الفقرات)؛ للعمل على تحسين قيم الثابت الضربي والثابت الجمعي نتيجة عدم دخول هذه الفقرة في حساب قيم هذه المتغيرات.

كما قام برنامج بحساب الأداء التفاضلي لاختبار العلوم، من خلال استخراج قيمة "مؤشر الأداء التفاضلي للاختبار DTF"، كما يظهر بالجدول (30).

الجدول (30)

معامل الارتباط بين العلامة الحقيقية للمجموعة المستهدفة والمجموعة المرجعية والأداء التفاضلي للاختبار

0.99964	Correlation between focal & reference true scores
0.00971	Differential test functioning (DTF)

حيث يلاحظ من الجدول أن الأداء التفاضلي للاختبار (0.00971) وهي قيمة أقل من القيمة الحرجة عند جميع مستويات الدلالة، وبالتالي لا يبدي الاختبار أداءً تفاضليًا، حيث تعبر هذه القيمة عن المجموع التراكمي لقيم مؤشر CDIF لجميع الفقرات الداخلة في الاختبار بناء على المعادلة (14) المشار لها في الفصل الأول من الدراسة، ولم تؤثر قيم "مؤشر CDIF" و"مؤشر DTF" على درجة نقاوة متغيرات المعادلة (الثابت الضربي والثابت الجمعي)، وهذا ما أثبتته قيم مؤشر إزالة الفقرة "2CDIF−NCDIF) item removal index"، حيث لم يتم إزالة أي فقرة من الاختبار، وبالتالي فالفقرات لا تبدي أداءً تفاضلياً بالاعتماد على مؤشرات CDIF ومؤشراتDTF)

تم إدخال البيانات مرة أخرى إلى برنامج (Equate v2.1)، لإجراء المرحلة الثانية من المعادلة وذلك بتكرار الخطوات السابقة بوضع جميع تقديرات معالم فقرات الجذع المشترك المتبقية التي لم تظهر أداءً تفاضليًا من المرحلة الأولى على مقياس مشترك (Common Metric) كما يتضح من الجدول (31)، لإعادة حساب معاملات المعادلة (الثابت الضربي والثابت الجمعي)؛ لاستخدامها لحساب مؤشرات (DFIT) النهائية. حيث بلغت قيمة الثابت الضربي (**0.9259**)، وهي قيمة أعلى من قيمة الثابت الضربي التي تم الحصول عليها من عملية المعادلة الأولية بمقدار (0.0032) بينما بلغت قيمة الثابت الجمعي (**0.097**)، وهي قيمة أقل من قيمة الثابت الجمعي التي تم الحصول عليها من عملية المعادلة الأولية بمقدار (0.0032)، وبالتالي تم تنقية قيم هذه المتغيرات من الأداء التفاضلي، وأصبحت هذه المعاملات أكثر استقرارًا، وقد اتضح ذلك من خلال قيمة معامل الارتباط بين العلامة الحقيقية للمجموعة المرجعية والمجموعة المستهدفة التي تم حسابها من خلال برنامج (Dfit v8.04)، والتي تدل

على مدى دقة المعادلة، حيث بلغت قيمة معامل الارتباط (0.99968)، وبالتالي هناك ارتفاع قليل على قيمة معامل الارتباط خاصة في الخانة العشرية الخامسة بمقدار (0.00004)، مما يدل على مدى التحسن بدقة المعادلة نتيجة استثناء الفقرات المشتركة التي أبدت أداءً تفاضلياً في المرحلة الأولية من المعادلة، ولم تدخل كفقرات مشتركة في المرحلة الثانية من المعادلة.

الجدول (31)

معالم الفقرات المقدرة للمجموعة المستهدفة بعد إجراء عملية المعادلة

c	a	b	رقم الفقرة
0.2976	0.9949	-0.0899	**1**
0.3106	0.5372	-0.1160	**2**
0.2995	0.6948	0.1809	**3**
0.1180	1.7401	2.9324	**4**
0.0821	1.3617	1.4808	**5**
0.2406	0.9463	0.6002	**7**
0.2208	2.4775	2.1262	**9**
0.2339	2.1724	0.8123	**10**
0.2756	1.0233	1.3266	**11**
0.2305	2.0166	0.2758	**12**
0.4925	2.2374	1.3364	**13**
0.2452	2.4330	1.8145	**14**
0.3370	1.5151	0.9006	**15**
0.2264	1.7221	1.6894	**18**
0.1580	2.2230	2.1052	**19**
0.2975	2.1573	1.1180	**20**
0.2860	2.4617	1.2949	**22**
0.1210	2.9441	0.5299	**23**
0.2946	2.7443	1.4459	**24**
0.1593	3.1554	2.0800	**25**
0.1926	1.4204	1.0630	**26**
0.2870	1.9713	0.6392	**27**
0.1931	1.4077	1.3296	**28**
0.2020	1.1609	3.5050	**29**
0.1770	1.1178	1.8269	**30**

يلاحظ من الجدول (31) أنه بعد وضع معالم الفقرات المقدرة من المجموعة المستهدفة (المحوسب) على مقياس المجموعة المرجعية (الورقي) من خلال عملية التحويل الخطي أن قيم معاملات الصعوبة تراوحت بين (–0.1160 إلى 3.5050)، أما معلمة التمييز فقد تراوحت بين (0.5372 إلى 3.1554)، في حين أن قيم التخمين تراوحت بين (0.0821 إلى 0.4925)، وبالتالي نكون قد ضمنا التكافؤ بين المجموعتين والدقة في تقدير هذه المعالم، وقد تبين هذا من خلال حساب معاملات المعادلة (الميل وثابت الانحدار)، وفي ضوء ذلك يلاحظ أن قيم معلمة

الصعوبة قد ارتفعت والتمييز قد انخفضت عما كانت عليه في مرحلة المعادلة الأولية في حين لم يحدث تغيير على قيم التخمين.

وباستخدام برنامج (Dfit v8.04) للمرة الثانية تم استخراج قيم مؤشر "الأداء التفاضلي غير التعويضي للفقرة (NCDIF)" للكشف عن الأداء التفاضلي في فقرات الاختبار، حيث تظهر الفقرة أداءً تفاضلياً لمتغير شكل تقديم الاختبار إذا كانت قيمة NCDIF المحسوبة أعلى من القيمة الحرجة، لدى قيام البرنامج بفحص الدلالة الإحصائية بحساب القيم الحرجة للأداء التفاضلي الخاصة بالاختبار وكل فقرة من فقراته والوسط الحسابي والانحراف المعياري للقيم الحرجة ضمن طريقة معلمة الفقرة التكرارية (Item Parameter Replication, IPR) عند مستويات دلالة مختلفة تتراوح بين (0.001 إلى 0.5)، علمًا بأن القيمة الحرجة للاختبار عبارة عن مجموع القيم الحرجة لــ NCDIF لجميع فقرات الاختبار.

الجدول (32)

النتائج النهائية للأداء التفاضلي لفقرات اختبار الرياضيات عبر المجموعتين المرجعية (ورقي) والمستهدفة (محوسب) من خلال قيم مؤشرات NCDIF، CDIF والتباين المشترك C(d,D) عند مستوى الدلالة المحدد

Sig.	NCDIF	CDIF	DIF C(d, D)	SD (d)	Mean (d)	رقم الفقرة
ns	0.00105	0.00163	0.002	0.026	-0.02	**1**
0.005	**0.00930**	0.00180	0.002	0.045	-0.086	**2**
ns	0.00019	0.00035	0	0.004	-0.013	**3**
ns	0.00019	-0.00001	0	0.006	0.012	**4**
0.05	**0.00690**	0.00072	0	0.028	0.078	**5**
ns	0.00169	0.00130	0.001	0.023	-0.034	**7**
ns	0.00206	0.00005	0	0.014	0.043	**9**
ns	0.00152	-0.00220	-0.002	0.037	-0.013	**10**
ns	0.00073	-0.00109	-0.001	0.022	0.015	**11**
ns	0.00027	-0.00071	-0.001	0.015	-0.007	**12**
0.01	**0.00890**	0.00846	0.008	0.094	-0.005	**13**
ns	0.00122	-0.00013	0	0.012	-0.033	**14**
ns	0.00132	0.00201	0.002	0.022	-0.029	**15**
ns	0.00205	-0.00226	-0.002	0.039	0.023	**18**
ns	0.00307	0.00002	0	0.005	0.055	**19**
ns	0.00161	0.00061	0.001	0.033	0.022	**20**
ns	0.00831	0.00489	0.005	0.075	0.051	**22**
ns	0.00365	-0.00304	-0.003	0.053	-0.028	**23**
ns	0.00255	0.00211	0.002	0.037	0.034	**24**
ns	0.00137	0.00027	0	0.034	-0.014	**25**
ns	0.00113	-0.00175	-0.002	0.027	-0.02	**26**
ns	0.00434	-0.00336	-0.003	0.041	-0.051	**27**
ns	0.00223	-0.00119	-0.001	0.03	-0.037	**28**
ns	0.00088	0.00050	0.001	0.028	-0.009	**29**
ns	0.00461	0.00031	0	0.014	0.066	**30**

الجدول (33)

قيم درجات القطع عند مختلف مستويات الدلالة α للفقرات والاختبار

SD	Mean	0.5	0.1	0.05	0.01	0.005	0.001	رقم الفقرة
0.00167	0.00175	0.00125	0.0038	0.00523	0.00833	0.0094	0.01185	**1**
0.00154	0.0016	**0.00112**	**0.00359**	**0.00461**	**0.00718**	**0.00925**	0.011	**2**
0.00153	0.00158	0.00112	0.00341	0.00442	0.00747	0.00856	0.01146	**3**
0.02737	0.00541	0.00125	0.00776	0.01466	0.06102	0.19897	0.45342	**4**
0.00182	0.00192	**0.00135**	**0.00406**	**0.00509**	0.0085	0.01024	0.01487	**5**
0.00142	0.00162	0.00118	0.00349	0.00463	0.00664	0.00729	0.00833	**7**
0.00672	0.00273	0.00143	0.0055	0.00785	0.0208	0.02587	0.05068	**9**
0.00431	0.00332	0.00196	0.00809	0.01067	0.02269	0.02426	0.03589	**10**
0.00281	0.00278	0.00185	0.00612	0.00861	0.01289	0.01582	0.01844	**11**
0.00195	0.00214	0.00157	0.0046	0.00592	0.00951	0.01086	0.01386	**12**
0.00183	0.00195	**0.00148**	**0.00409**	**0.00539**	**0.00837**	0.00966	0.01122	**13**
0.0027	0.0024	0.00157	0.00534	0.00715	0.01167	0.01666	0.02278	**14**
0.00201	0.00225	0.0017	0.00477	0.00601	0.00947	0.0109	0.01501	**15**
0.00295	0.00224	0.00143	0.00506	0.00701	0.01264	0.01612	0.03197	**18**
0.00421	0.0026	0.00152	0.0055	0.00766	0.01575	0.02848	0.05161	**19**
0.00315	0.00292	0.00191	0.00688	0.0089	0.015	0.01557	0.0267	**20**
0.00331	0.0028	0.00157	0.00667	0.00906	0.0158	0.01932	0.0223	**22**
0.00281	0.00248	0.00151	0.00575	0.00829	0.01291	0.01443	0.01909	**23**
0.00352	0.00289	0.00177	0.00644	0.00937	0.01553	0.02211	0.03453	**24**
0.00845	0.0026	0.00126	0.00437	0.00654	0.02542	0.04215	0.12354	**25**
0.00317	0.00286	0.00175	0.00703	0.00895	0.01457	0.01839	0.02206	**26**
0.00354	0.00308	0.00184	0.00707	0.00974	0.01838	0.01977	0.02458	**27**
0.0026	0.00248	0.00163	0.00537	0.00739	0.01246	0.01439	0.02038	**28**
0.00695	0.0032	0.00156	0.00641	0.01007	0.0287	0.04289	0.0837	**29**
0.00182	0.00185	0.0013	0.00396	0.00515	0.00894	0.01102	0.01277	**30**
0.0632	0.06398	0.04752	0.12947	0.17216	0.33553	0.36149	0.48108	**DTF**

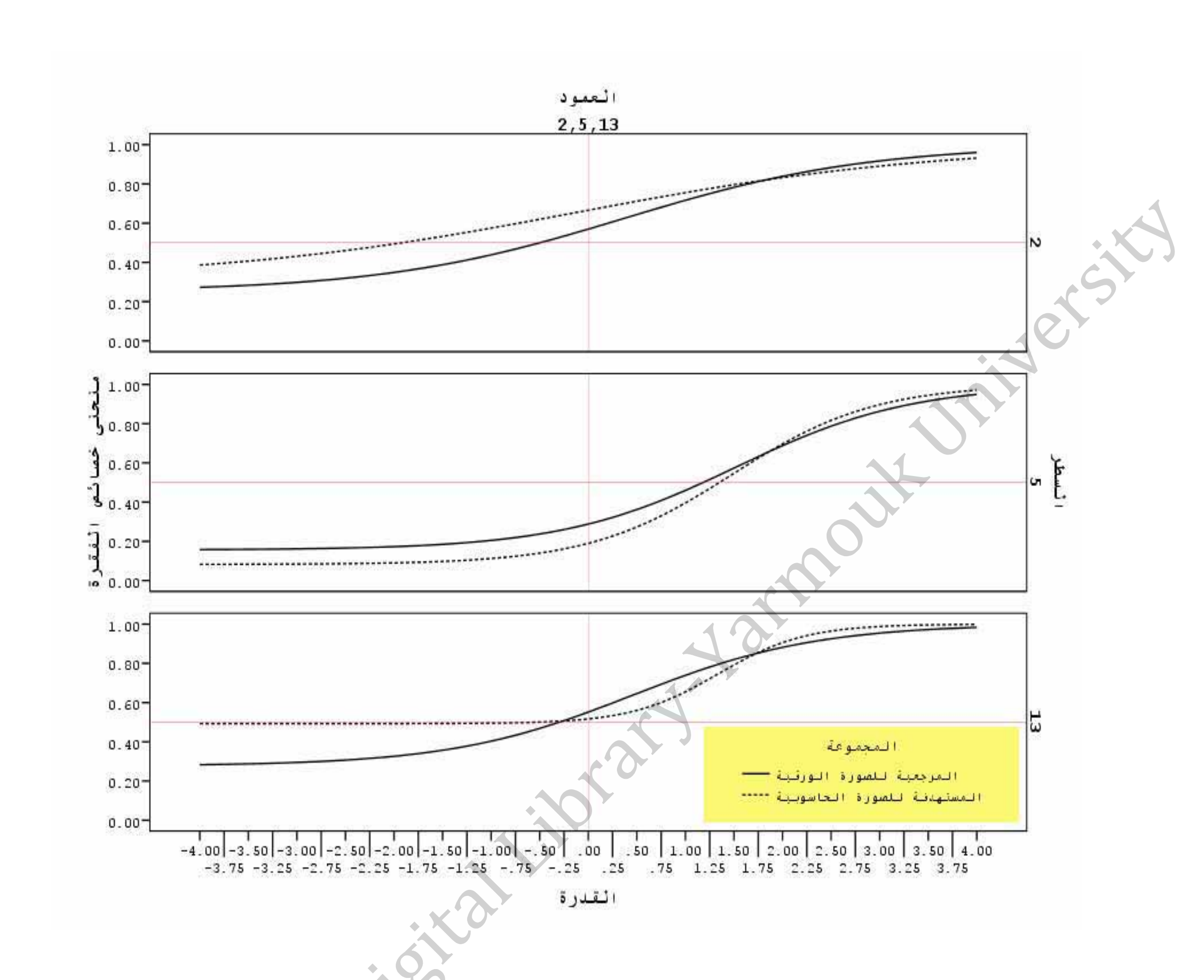
في ضوء قيم مؤشر NCDIF، استمرت الفقرات رقم (2 ,5 ,13) في إبداء الأداء التفاضلي تبعا شكل تقديم الاختبار (ورقي، محوسب) بنسبة 12%من فقرات الاختبار المكون من (25) فقرة بصورته النهائية، بالرغم من إزالة اثرهنّ من الفقرات المشتركة في المرحلة الأولية من المعادلة، حيث بلغت قيم هذا المؤشر (0.0093, 0.0069 ,0.0089) على التوالي، وهي قيم أعلى من القيمة الحرجة عند مستوى الدلالة (α=0.01, 0.05, 0.005) على التوالي، لدى فحص الدلالة الإحصائية بطريقة IPR.

وتتعلق الفقرة 2 بمحور الأحياء، وتهدف إلى معرفة نمط حياة بعض الكائنات الحية بناء على معلومات معطاة، حيث تتطلب معرفة نوع الانعزال التكاثري ما بين الحية ذات الطوق والحية السامة. أما الفقرة 5 فهي فقرة تتعلق بمحور الأحياء، وتهدف إلى استقصاء أثر تركيز

الأكسجين على نمو بعض الكائنات الحية، وتتطلب معرفة نوع البكتيريا في الشكل حسب حاجتها للأكسجين، في حين تتعلق الفقرة 13 بمحور الفيزياء، وتهدف إلى حساب الجهد الكهربائي في مصباح كهربائي، حيث تتطلب حساب قيمة الجهد الكهربائي في مصباح يستهلك طاقة معلومة عندما يمر فيه تيار معلوم القيمة.

كما تبين أن (22) فقرة من أصل 25 فقرة لم تظهر أداءً تفاضليًا دال إحصائيًا لمتغير شكل تقديم الاختبار بنسبة 88% تقريباً، حيث كانت قيمة مؤشر NCDIF أقل من قيمة القطع عند جميع مستويات الدلالة (α)، لدى فحص الدلالة الإحصائية بطريقة (IPR)، حيث أن قيمة مؤشر NCDIF للمجموعتين تكاد لا تؤخذ بعين الاعتبار مقارنة بالقيمة الحرجة (درجة القطع) التي تم اعتمادها معيارًا للحكم على وجود الأداء أو عدم وجوده، ويتضح ذلك في الفقرات (,29 ,30 1, 3, 4, 7, 9, 10, 11, 12, 14, 15, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28).

كما تم التأكد من مدى إظهار الفقرات للأداء التفاضلي من خلال رسم منحنيات خصائص الفقرة للمجموعتين المستهدفة والمرجعية باستخدام البرمجية الإحصائية (SPSS) للنموذج الثلاثي المعلمة؛ للاستدلال على الأداء التفاضلي للفقرة، ولبيان الاختلاف بين منحنيات خصائص الفقرة للمجموعتين (المستهدفة والمرجعية) على اختبار العلوم على مستوى الفقرة الواحدة، وبيان نوع الأداء التفاضلي فيما إذا كان منتظماً أو غير منتظم، كما يتضح في الشكل (6).



شكل (6): منحنى خصائص الفقرة للفقرات (2 ,5 ,13) التي أظهرت أداء تفاضليا تبعا لشكل تقديم الاختبار

تبين من خلال تفحص منحنى خصائص الفقرة للفقرات (2 ,5 ,13) أن الفقرة (2) تبدي أداءً تفاضلياً غير منتظم، حيث أن احتمال الإجابة الصحيحة عن الفقرة يختلف تبعًا لطريقة تقديم الاختبار عند المفحوصين المتساوين في القدرة، وأن منحنى المجموعة المستهدفة أقل تمييزًا من منحنى المجموعة المرجعية، فاحتمال الإجابة الصحيحة عن الفقرة أعلى للمجموعة المستهدفة (المحوسب) من المجموعة المرجعية (الورقي) ضمن مستويات القدرة المتدنية والمتوسطة؛ أي أن الفقرة تعمل لصالح المجموعة المستهدفة (المحوسب) عند هذه المستويات من القدرة، واحتمال الإجابة الصحيحة أقل للمجموعة المستهدفة من المجموعة المرجعية ضمن مستويات

القدرة المرتفعة؛ أي أن الفقرة تعمل لصالح المجموعة المرجعية (الورقي) عند هذا المستوى من القدرة.

كما أظهرت الفقرة (5) أداءً تفاضلياً غير منتظم تبعاً لطريقة تقديم الاختبار عند المفحوصين المتساوين في القدرة، فاحتمال الإجابة الصحيحة على الفقرة: أعلى للمجموعة المستهدفة (المحوسب) من المجموعة المرجعية (الورقي) ضمن مستويات القدرة المرتفعة؛ أي أن الفقرة تعمل لصالح المجموعة المستهدفة (المحوسب) عند هذا المستوى من القدرة، واحتمال الإجابة الصحيحة أقل للمجموعة المستهدفة من المجموعة المرجعية ضمن مستويات القدرة المنخفضة والمتوسطة؛ أي أن الفقرة تعمل لصالح المجموعة المرجعية (الورقي) عند هذه المستويات من القدرة.

وكذلك الأمر بالنسبة للفقرة (13) التي أظهرت أداءً تفاضلياً غير منتظم تبعاً لطريقة تقديم الاختبار عند المفحوصين المتساوين في القدرة، فاحتمال الإجابة الصحيحة عن الفقرة أعلى للمجموعة المستهدفة (المحوسب) من المجموعة المرجعية (الورقي) ضمن مستويات القدرة المرتفعة والمنخفضة؛ أي أن الفقرة تعمل لصالح المجموعة المستهدفة (المحوسب) عند هذه المستويات من القدرة، واحتمال الإجابة الصحيحة أقل للمجموعة المستهدفة من المجموعة المرجعية ضمن مستوى القدرة المتوسط؛ أي أن الفقرة تعمل لصالح المجموعة المرجعية (الورقي) عند هذا المستوى من القدرة.

وعند تفحص منحنيات خصائص الفقرة للفقرات التي لم تبد أداءً تفاضليًا نجد أن منحنيات خصائص الفقرة لكل فقرة من هذه الفقرات عبر المجموعتين المرجعية والمستهدفة تعكس مقداراً قليلاً من عدم الانتظام في منحنياتها إذا ما قورنت مع الفقرات التي تظهر منحنياتها أداء تفاضلياً،

حيث أن أداء المفحوصين عليها من كلا المجموعتين المرجعية والمستهدفة كان متقاربــاً، أو أن الفرق بين المنحنيين للمجموعتين كان قريباً من الصفر، بحيث يكاد ينطبــق المنحنيــان علــى بعضهما عند جميع مستويات القدرة المختلفة، وبالتالي فالاختلاف بين المنحنيين لا يتعدى معياراً معيناً وهو خطأ المعاينة.(انظر الملحق 10)

كما تم استخراج قيم مؤشر الأداء التفاضلي التعويضي للفقرة CDIF، لمعرفة العلاقة بين الأداء التفاضلي للفقرة والأداء التفاضلي للاختبار، فعند فحص قيم مؤشر الأداء التفاضلي التعويضي للفقرة CDIF من الجدول (32) يلاحظ أن بعض قيمها سالب. كما في الفقرات (28, 4, 10, 11, 12, 14, 18, 23, 26, 27)، والأخر موجب كما في الفقرات (24, 25, 29, 30, 1, 2, 3, 5, 7, 9, 13, 15, 19, 20, 22)، وأن أعلى قيمة لها (0.00846) للفقرة (13)، وأقل قيمة لها (0.00336-) للفقرة (27)، وبالتالي تعمل القيم السالبة والموجبة للمؤشر بإلغاء بعضها بعضاً عندما تجمع على مستوى الاختبار (Cancelation Effect)، وما يتبقى يعكس مدى إسهام الفقرات في الأداء التفاضلي للاختبار، فعند الكشف عن الأداء التفاضلي للاختبار من خلال استخراج قيمة مؤشر الأداء التفاضلي للاختبار DTF كما يتضح في الجدول (34)

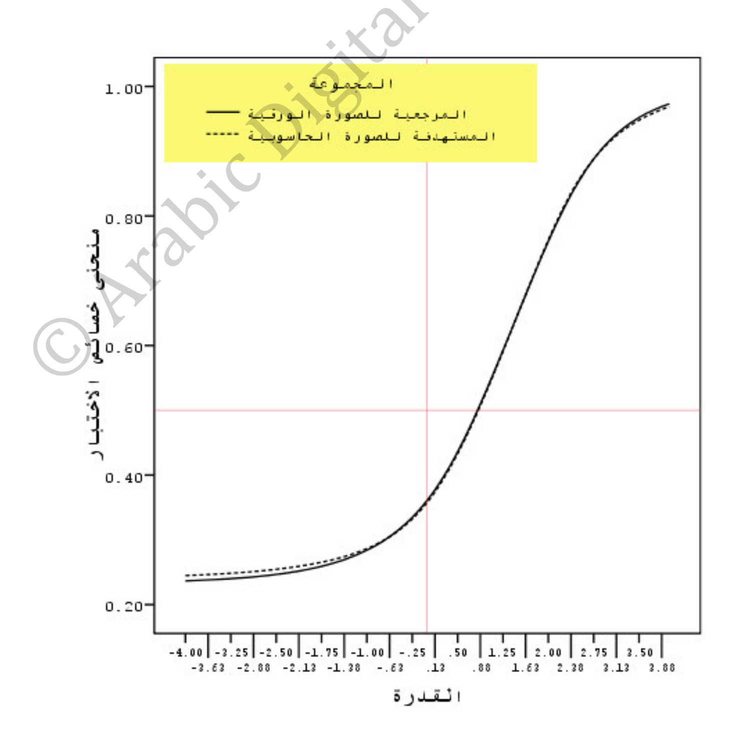
جدول (34)

معامل الارتباط بين العلامة الحقيقية للمجموعة المستهدفة والمجموعة المرجعية، والأداء التفاضلي للاختبار

0.99968	Correlation between focal & reference true scores
0.00929	Differential test functioning (DTF)

تبين أن الأداء التفاضلي للاختبار قد بلغ (0.00929) وهي قيمة أقل من القيمة الحرجــة عند جميع مستويات الدلالة (α) (القيمة الحرجة للاختبار عبارة عن مجمــوع القــيم الحرجــة لمؤشر NCDIF لجميع فقرات الاختبار)، وهذه القيمة تعبر عن المجموع التراكمي لقيم مؤشــر

CDIF لجميع الفقرات الداخلة في الاختبار؛ أي أن الاختبار لا يبدي أداءً تفاضليًا دال إحصـــائيًا وبالتالي فإن التأثير التراكمي للأداء التفاضلي للفقرات على الاختبار كان قليل جدًا، وقد تم تأكيد ذلك بعمل فحص لمؤشر إزالة الفقرة item removal index (2CDIF-NCDIF) بدءً بالفقرات الأعلى مؤشر إزالة لتحديد الفقرات التي ستزال من الاختبار (الفقرات التي تزال توصف بأنهـا تبدي أداءً تفاضلياً بالاعتماد على مؤشرات CDIF ومؤشراتDTF) أي أن مدى إسهام الفرق في احتمال الإجابة الصحيحة على الفقرة عند مستوى قدرة معين في الفرق فـي احتمـال الإجابـة الصحيحة على الاختبار عند نفس مستوى القدرة لم يرقَ إلى مستوى الدلالة الإحصائية، لجعـل الاختبار يظهر أداءً تفاضليًا، وأن درجة ارتباط الفقرات مع بعضها كبير وبالتالي هناك اتسـاق بين ما تقيسه الفقرة وبين ما يقيسه الاختبار وهذا ما أكده منحنى خصائص الاختبار لدى رسمه باستخدام برنامج (SPSS)، كما يظهر الشكل (7).



شكل (7): منحنى خصائص اختبار العلوم للمجموعتين المستهدفة والمرجعية.

الفصل الخامس

مناقشة النتائج والتوصيات

يتناول هذا الفصل خلاصة نتائج الدراسة، ومناقشتها، وتفسيرها، لكل سؤال على انفـــراد، كما يتناول عرضًا للتوصيات والمقترحات التي توصي بها الدراسة في ضوء النتائج التـــي تـــم التوصل إليها.

أولاً: مناقشة النتائج المتعلقة بالسؤال الأول: "**هل تظهر فقرات اختبار الرياضيات الوطني لضبط نوعية التعليم لطلبة الصف العاشر (2011/2010) أداءً تفاضلياً ذا دلالة إحصائية يعود إلى طريقة تقديم الاختبار (ورقي مرجعي، محوسب مستهدف)؟**".

أظهرت النتائج التي توصلت لها الدراسة الحالية فيما يتعلق بالأداء التفاضلي لفقرات اختبار الرياضيات الوطني لضبط نوعية التعليم تبعًا لمتغير طريقة تقديم الاختبار (ورقي، محوسب)، وجود أداءً تفاضليًا لمتغير طريقة تقديم الاختبار في فقرتين من فقرات الاختبار البالغ عددها (24)، وذلك بنسبة (8.33%) إحداهما أبدت أداءً تفاضليًا منتظم لصالح الشكل المحوسب للاختبار عند مختلف مستويات القدرة بنسبة (4.17%) تقريبًا، وهي الفقرة 28 التي تتعلق بمحور الإحصاء والاحتمالات وهذه النتيجة لا تتفق مع نتيجة دراسة فيسر (Vesser, 1998) التي توصلت من خلالها إلى أن نتائج الطلبة كانت أفضل بواسطة الطريقة التقليدية منها في القياس التكيفي أو الاختبارات المبرمجة مع أن الخصائص السيكومترية (الثبات والصدق) متماثلة في الاختبارات الثلاثة. كما لا تتفق كذلك مع دراسة بينيت وبارسول واورانج وساندن وكابلان ويان (Bennett, Braswell, Oranje, Sandene, Kaplan & Yan, 2008)، التي

تم التوصل من خلالها إلى أن الاختبار المقدم بشكل محوسب كان أصعب بشكل دال إحصائيًا من الاختبار الورقي. إلا أنها تتفق مع نتيجة دراسة جو ودراك وولف (Gu, Drake & Wolfe, 2006) التي تم التوصل من خلالها إلى أن 38% من الفقرات أظهرت أداء تفاضليًا لصالح الشكل الالكتروني، وأنه بالرغم من أن النتائج أظهرت أن الاختلاف في تنسيق الصفحة، وكيفية الإجابة عن الفقرة على الشكل الالكتروني كانت مسئوله بشكل قليل عن الأداء التفاضلي على فقرات هذا الشكل، إلا أن الاختلافات في الملاحظات الرياضية الواردة في نص الفقرة، والاختلاف في المحتوى الرياضي للفقرات كان لها علاقة قوية مع الأداء التفاضلي على فقرات الشكل الالكتروني. ويتفق كذلك مع دراسة بوهن وبوتن وكيم (Puhan, Boughton & Kim, 2007) التي تم التوصل من خلالها إلى أن هناك ثلاثة فقرات أظهرت أداء تفاضليًا لصالح الشكل المحوسب فشلت المراجعة الموضوعية من قبل المحكمين في تحديد فيما إذا كانت الفروق في الأداء يمكن تفسيرها نتيجة الاختلاف في شكل الاختبار (ورقي أو الكتروني)، لذلك لم يتم تحديد الأسباب المسئولة عن الأداء التفاضلي.

في حين أن الفقرة 13 أبدت أداءً تفاضليًا غير منتظم لصالح الشكل المحوسب للاختبار ضمن مستويات القدرة الدنيا تقريبًا، ولصالح الشكل الورقي للاختبار ضمن مستويات القدرة المتوسطة والعليا تقريبًا بنسبة (4.17%) تقريبًا، وهي فقرة تتعلق بمحور الجبر وهذا يتفق مع دراسة شوارتز وريتش وبودرابسكي (Schwarz, Rich & Podrabsky, 2003) التي توصل من خلالها إلى وجود فقرتين أظهرتا أداءً تفاضلياً على اختبار (TABE) حيث كانت الفروق بين المجموعات على الاختبار المحوسب والورقي أكثر وضوحًا في الجزء السفلي من توزيع القدرة.

وقد عزت الدراسة السبب في ظهور هذه النتيجة إلى انخفاض قيمة معامل الثبات لدرجات الاختبار على الشكل الورقي إذا ما قورن مع الشكل المحوسب، ويعود هذا إلـــى ارتفـــاع قـــيم

معاملات الصعوبة لهاتين الفقرتين الناتج عن أن طبيعة مضمونهما يركز على بعض المهارات المتعلقة بالجبر، والإحصاء والاحتمالات، والتي تتطلب من المفحوص مهارات عقلية دنيا (تذكر، فهم، تطبيق) وهذا يجعل الكفة ترجح لصالح الاختبار المحوسب؛ بسبب سلاسة التعامل مع الاختبار باستفادة المفحوصين من مزايا الاختبارات المحوسبة خاصة المتعلقة بتنسيق الصفحة بإمكانية تغير حجم الخط بالشكل المحوسب، وعدم إمكانية ذلك بالشكل الورقي، والمتعلقة بسرعة الاستجابة، فالاستجابة تكون أسرع بالشكل المحوسب من خلال النقر بواسطة الفأرة، أما الاستجابة بالشكل التقليدي فتتم من خلال تظليل الدائرة بالكامل بقلم رصاص ليتيقن الطالب من إمكانية قراءتها من القارئ الضوئي، وأيضًا السرعة في عمليات الحذف والتعديل بسبب استخدام تقنية النقر بواسطة الفأرة، بينما يحتاج ذلك إلى استخدام الممحاة وإعادة التظليل مرة أخرى بالشكل الورقي للاختبار، مما قد يأخذ وقت أطول على حساب إجابة باقي فقرات الاختبار مما قد يؤثر على مقدار تركيز المفحوص للاستدلال إلى الإجابة الصحيحة الأمر الذي يزيد من الإرباك (العبء المعرفي) لدى المفحوص، وهذا الإرباك يحد من استرسال الاستدلال للوصول للإجابة الصحيحة وهذا يجعل الطالب يلجأ إلى التخمين العشوائي ويتضح هذا من خلال قيم التخمين عبر شكلي الاختبار، حيث يلاحظ ارتفاع قيمة التخمين على الشكل الورقي للاختبار لكلا الفقرتين اللتين أظهرتا أداءً تفاضليًا، إذا ما قورنت مع قيمته على الشكل المحوسب، والناتج عن ارتفاع معاملات الصعوبة لهذه الفقرات ويتضح ذلك أيضًا من خلال استقراء منحنى خصائص الفقرة لهاتين الفقرتين، فبالرغم من أن نوع الأداء التفاضلي للفقرة رقم 13 غير منتظم إلا أن الملاحظ أن الفرق الأكبر في الأداء لصالح الشكل المحوسب أكثر وضوحاً عند ذوي القدرة المتدنية مما يعني استفادتهم من مزايا الاختبار المحوسب أكثر من غيرهم من المفحوصين، وفيما يتعلق بالفقرة رقم 28 التي تبدي أداءً تفاضليًا منتظم، فإن المفحوصين استفادوا من مزايا الاختبار المحوسب عند جميع مستويات القدرة، وهذا يتفق مع

دراسة جونسون وغرين (Johnson & Green, 2006)، الذين توصلا إلى أن التحليل الـدقيق للبيانات أشار إلى أن نوع السؤال، طريقة طرح السؤال، والأرقام التي يحتويها قد تتفاعل مـع طريقة عرض الاختبار في التأثير على مدى استعداد المفحوصين لإظهـار خطـوات حلهـم، والاستراتيجيات المتبعة في الحل، كما أشارت النتائج إلى أن أنواع معينة من الأسئلة في بعض المجالات قد يكون لها تأثيرات مختلفة حسب شكل تقديم الاختبار، و يتفق مع دراسة بـومريتش وبوردن (Pommerich, Burden, 2000) التي أشارا فيها إلى أن هناك العديد من العوامـل التي تتفاعل مع قدرة المفحوصين على الاستجابة على فقرات الاختبار، إضافةً إلـى محتـوى الفقرة ومن هذه العوامل: فواصل الصفحات والأسطر، ميزات الفقرات، التظليـل، خصـائص الفقرة، بالإضافة إلى عوامل الحركة مثل: مراجعة الفقرات، معاينة الفقرات، قـدرات الحـذف للإجابات الخاطئة.

الفقرات التي أبدت أداءً تفاضليًا على مستوى الفقرة كانت قليلة حيث كانت النسبة (8.33%) من الاختبار ككل، وبالتالي فإنه عمليًا لن يكون هناك ظلم كبير بين المجموعتين تعود إلى شكل تقديم الاختبار وهي نتيجة مهمة لمطوري الاختبار يمكن استخدامها بدقة لتقييم مدى امتلاك طلاب الصف العاشر لمهارات التعلم الأساسية في الرياضيات، واتخاذ قرارات مهمة عن واقع التعليم في الأردن، وعن مدى ملائمة المناهج الأردنية لقدرات الطلبة، مما قد يعكس مدى التحسن في مستوى الأداء الذي يحصل نتيجة التجديدات التربوية ودمج التكنولوجيا بالعملية التعليمية في إطار المرحلة الثانية من مشروع التطوير التربوي لاقتصاد المعرفة.

كما أن غالبية فقرات اختبار الرياضيات خالية من الأداء التفاضلي حيث أن 22 فقرة من فقرات الاختبار لم تبد أداءً تفاضليًا تبعًا لمتغير طريقة تقديم الاختبار بنسبة 91.67% تقريبًا

وبالتالي فمن الممكن الاعتماد عليها مستقبلاً في تطوير اختبارات تخلو من الأداء التفاضلي لمتغير طريقة تقديم الاختبار إلى حد معقول.

لم يبدِ الاختبار ككل أداءً تفاضليًا حيث بلغت قيمة مؤشر DTF (0.04419)؛ أي أن مدى إسهام الفرق في احتمال الإجابة الصحيحة على الفقرة عند مستوى قدرة معين في الفرق في احتمال الإجابة الصحيحة على الاختبار عند نفس مستوى القدرة لم يرق إلى مستوى الدلالة الإحصائية لجعل الاختبار يظهر أداءً تفاضليًا وبالتالي هناك اتساق بين ما تقيسه الفقرة وبين ما يقيسه الاختبار وهذا يعني ضمنيًا أمرين؛ الأوّل منهما تأكيد افتراض أحادية البعد للاختبار والآخر منهما تحقق خاصية اللاتغاير، وهذا ما أشارت له دراسة المومني (2012) التي توصلت إلى عدم وجود تغير في فترات الثقة حول معالم الاختبار، وفق نموذج راش، يعزى لاختلاف الطريقة التي تم فيها تقديم الاختبار، وبالتالي فإن تحقق خاصية اللاتغاير تعني ضمنيًا عدم وجود أداء تفاضلي يعود لشكل تقديم الاختبار، كما يتفق مع دراسة بوهن وبوتن وكيم (Puhan, Boughton & Kim, 2007) التي توصلت إلى أن قيمة حجم الأثر غير دالة إحصائيًا، حيث كانت قليلة ($d < 0.20$) مما يشير إلى عدم وجود فروق كبيرة في الأداء على نسختي الاختبار، وتتفق هذه النتيجة مع نتيجة دراسة بودمان وروبنسون (Bodmann & Robinson, 2004) التي توصلا فيها إلى أنه لا يوجد أثر لشكل تقديم الاختبار على درجات المفحوصين بينما كان هنالك أثر على زمن إنهاء الاختبار حيث أنهى الطلاب الاختبار على النسخة الإلكترونية بسرعة أكبر من زمن إنهاء الطلاب على النسخة الورقية. وكذلك مع دراسة هوركاي وبينت والين وكابلان ويان (Horkay, Bennett, Allen, Kaplan, & 2006

Yan,) التي تم التوصل فيها إلى عدم وجود فروق ذات دلالة إحصائية بين أداء المشاركين على الاختبار المحوسب وأدائهم على الاختبار الورقي.

ثانيًا: مناقشة النتائج المتعلقة بالسؤال الثاني: "**هل تظهر فقرات اختبار العلوم الوطني لضبط نوعية التعليم لطلبة الصف العاشر (2011/2010) أداءً تفاضليًا ذا دلالة إحصائية يعود إلى طريقة تقديم الاختبار (ورقي مرجعي، محسوب مستهدف)**؟".

أظهرت النتائج التي توصلت لها الدراسة الحالية، فيما يتعلق بالأداء التفاضلي لفقرات اختبار العلوم الوطني لضبط نوعية التعليم، وجود أداءً تفاضليًا لمتغير طريقة تقديم الاختبار في ثلاث فقرات من فقرات الاختبار البالغ عددها (25) فقرة، وذلك بنسبة (12%) جميعها أبدت أداءً تفاضليًا غير منتظم الفقرة 2 كانت لصالح الشكل المحوسب ضمن مستوى القدرة الدنيا، ولصالح الشكل الورقي ضمن مستوى القدرة المرتفع، الفقرة 5 لصالح الشكل المحوسب ضمن مستوى القدرة العليا، ولصالح الشكل الورقي ضمن مستوى القدرة المنخفض والمتوسط، الفقرة 13 لصالح الشكل المحوسب ضمن مستوى القدرة الدنيا والعليا، ولصالح الشكل الورقي ضمن مستوى القدرة المتوسط، في ضوء هذه النتائج، يلاحظ أن هناك تباين في اتجاه الأداء التفاضلي، حيث يكون تارةً لصالح الشكل الورقي، وتارةً أخرى لصالح الشكل المحوسب.

عزت الدراسة السبب في ظهور هذه النتيجة إلى انخفاض قيمة معامل الثبات لدرجات الاختبار على الشكل الورقي إذا ما قورن مع الشكل المحوسب والناتج عن أن طبيعة مضمون هذه الفقرات يركز على بعض المهارات المتعلقة بالأحياء، والفيزياء، حيث تتطلب الفقرتين (2 ,13) من المفحوص مهارات عقلية دنيا (تذكر، فهم، تطبيق) وهذا يجعل الكفة ترجح لصالح الاختبار المحوسب بسبب سلاسة التعامل مع الاختبار باستفادة المفحوصين من مزايا الاختبارات

المحوسبة، خاصة المتعلقة بتنسيق الصفحة بإمكانية تغير حجم الخط بالشكل المحوسب، وعـــدم إمكانية ذلك بالشكل الورقي، والمتعلقة بسرعة الاستجابة، فالاستجابة تكـــون أســـرع بالشـــكل المحوسب من خلال النقر بواسطة الفأرة، أما الاستجابة بالشكل التقليدي فتتم من خلال تظليـــل الدائرة بالكامل بقلم رصاص ليتيقن الطالب من إمكانية قراءتها من القارئ الضـــوئي، وأيضًـــا السرعة في عمليات الحذف والتعديل بسبب استخدام تقنية النقر بواسطة الفأرة، بينما يحتاج ذلك إلى استخدام الممحاة وإعادة التظليل مرة أخرى بالشكل الورقي للاختبار، مما قد يأخـــذ وقـــت أطول على حساب إجابة باقي فقرات الاختبار وهذا يؤثر على مقدار تركيز المفحوص للاستدلال إلى الإجابة الصحيحة الأمر الذي يزيد من الإرباك (العبء المعرفي) لدى المفحـــوص، وهـــذا الإرباك يحد من استرسال الاستدلال للوصول للإجابة الصحيحة، مما يجعل الطالب يلجـــأ إلـــى التخمين العشوائي، وعند التبصر في مضمون الفقرة رقم 5 وجد أنها تتطلب مـــن المفحـــوص مهارات عقلية عليا (تحليل، تركيب، تقويم) مما يجعل الكفة ترجح لصالح الاختبار الورقي؛ لأن الفقرة تستوجب تآزرًا حركيًا بصريًا أثناء التفاعل معها ولا يتأتى ذلـــك إلا مـــن خـــلال القلـــم والورقة؛ لما تستوجبه من وضع تدوينات خاصة بالمستجيب على الورقة البيضاء التي يزود بها مما يرفع من شدة التركيز في الاستدلال إلى الإجابة الصحيحة، وهذا ما يتعذر تحقيقـــه علـــى الشكل المحوسب؛ لان التآزر البصري الحركي في تلك الحالة يكون ما بين الفأرة والشّاشة دون مداخلات جانبية تمكن المستجيب من رفع شدة التركيز في الاستدلال على الرغم مـــن تزويـــد المفحوص بورقة بيضاء في معرض إجابته للاختبار المحوسب، الأمر الذي قد يزيد من الإرباك لديه مما يحد من استرسال الاستدلال للوصول إلى الإجابة الصحيحة في حـــال تطلـــب الفقـــرة مهارات عقلية عليا، ويتضح هذا من خلال قيم التخمين عبر شكلي الاختبار، حيث يلاحظ ارتفاع قيمة التخمين على الشكل الورقي للاختبار للفقرتين (2 ,13) اللتين أظهرتا أداءً تفاضليًا إذا مـــا

قورنت مع قيمته على الشكل المحوسب، ويتضح ذلك أيضًا من خلال استقراء منحنى خصائص الفقرة لهاتين الفقرتين، فبالرغم من أن نوع الأداء التفاضلي غيــر منــتظم إلا أن الملاحــظ أن المساحة الأكبر للفرق في الأداء لصالح الشكل المحوسب أكثر وضوحاً عند ذوي القدرة المتدنية مما يعني استفادتهم من مزايا الاختبار المحوسب أكثر من غيرهم من المفحوصين، وفيما يتعلق بالفقرة رقم 5 يلاحظ ارتفاع قيمة التخمين على الشكل المحوسب للاختبار، ومن استقراء منحنى خصائص الفقرة وجد أن المساحة الأكبر للفرق في الأداء لصالح الشكل الورقي أكثر وضــوحاً عند ذوي القدرة المتدنية، وبما أن من أهداف الدراسة التوسع في تطبيق الاختبارات المحوســبة فإنه قد يكون من المفضل إعادة صياغة الفقرات بشكل يحد من الإرباك لدى المفحوص، بإضافة مميزات للفقرة كأن تكون بعض فقرات الاختبار من نوع الفقرات التي تراعي المعرفة الجزئية، وهذا يتفق مع دراسة بومريتش وبوردن (Pommerich, Burden, 2000) التي أشار فيها إلى أن هناك العديد من العوامل التي تتفاعل مع قدرة المفحوصين للإجابة عن فقــرات الاختبــار، إضافةً إلى محتوى الفقرة ومن هذه العوامل: فواصل الصفحات والأســطر، ميــزات تصــميم الفقرات، التظليل، خصائص الفقرة، بالإضافة إلى عوامل الحركة مثل: مراجعة الفقرات، معاينة الفقرات، قدرات الحذف للإجابات الخاطئة.

كما أن خصائص المفحوصين قد تسهم بالعديد من الآثار الناتجة عن شكل تقديم الاختبار وخاصة عدم جدية الممتحنين التي قد تلعب دور في ذلك خاصة أن دافعية المفحوصين أكبر على الشكل المحوسب منه على الشكل الورقي كما أن إنهاء الفقرات بزمن أقل في الشكل المحوسب يجعل هناك جدية أكثر وهذا يتفق مع دراسة بياو (Piaw, 2010) التي توصل من خلالها إلى أن شكل الاختبار يؤثر بشكل دال إحصائيًا على زمن الاختبار والدافعية ويعمل على تقليل زمن الاختبار وزيادة الدافعية للمفحوصين عبر الشكل المحوسب للاختبار.

لم يبد الاختبار ككل أداءً تفاضليًا وهذا يعني ضمنيًا أمرين؛ الأوّل منهما تأكيد افتراض أحادية البعد للاختبار والآخر منهما تحقق خاصية اللاتغير وهذا ما أشارت له دراسة المومني (2012) التي توصلت إلى عدم وجود تغير في فترات الثقة حول معالم الاختبار، وفق نموذج راش، يعزى لاختلاف الطريقة التي تم فيها تقديم الاختبار، وبالتالي فإن تحقق خاصية اللاتغير تعني ضمنيًا عدم وجود أداء تفاضلي يعود لشكل تقديم الاختبار، ويتفق هذا أيضًا مع دراسة كيم وهيونه (Kim, Huynh, 2010) حيث أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين الاختبار باستخدام الحاسوب والاختبار بالقلم والورقة.

كما أن معظم فقرات اختبار العلوم خالية من الأداء التفاضلي حيث أن 22 فقرة من فقرات الاختبار لم تبد أداءً تفاضليًا تبعًا لمتغير شكل تقديم الاختبار بنسبة 88% وبالتالي من الممكن الاعتماد عليها مستقبلاً، في تطوير اختبارات تخلو من الأداء التفاضلي لمتغير شكل تقديم الاختبار إلى حد معقول.

الفقرات التي أبدت أداءً تفاضليًا على مستوى الفقرة كانت قليلة، حيث كانت النسبة (12%) من الاختبار ككل، وبالتالي فإنه عمليًا لن يكون هناك ظلم كبير بين المجموعتين تعود إلى طريقة تقديم الاختبار، وهو نتيجة مهمة لمطوري الاختبار، يمكن استخدامها بدقة لتقييم مدى امتلاك طلاب الصف العاشر لمهارات التعلم الأساسية في العلوم واتخاذ قرارات مهمة عن واقع التعليم في الأردن وعن مدى ملائمة المناهج الأردنية لقدرات الطلبة مما قد يعكس مدى التحسن في مستوى الأداء الذي يحصل نتيجة التجديدات التربوية ودمج التكنولوجيا بالعملية التعليمية في إطار المرحلة الثانية من مشروع التطوير التربوي لاقتصاد المعرفة.

التوصيات:

توصي الدراسة بما يلي:

1. إجراء المزيد من الدراسات القائمة على تحري الأداء التفاضلي على الاختبار الوطني لضبط نوعية التعليم بحيث تكون أكثر شمولية بأن تغطي المجالات الأربعة للاختبار (علوم، رياضيات، لغة عربية، لغة إنجليزية).

2. إجراء دراسة لتحري الأداء التفاضلي لشكل تقديم الاختبار على الاختبار الوطني عبر الصفوف الدراسية الرابع، الثامن، والعاشر والعمل على المقارنة بينها وفق نظرية الاستجابة للفقرة.

3. توصية لوزارة التربية والتعليم بالتوسع في تطبيق الامتحانات العامة حاسوبيًا.

المراجع والمصادر

أولاً: المراجع العربية

حمادنة، إياد. (2007). *الأداء التفاضلي لفقرات اختبارات "تحديد الكفاءة اللغوية في اللغة الانجليزية "في الجامعات الأردنية : دراسة مقارنة*. رسالة دكتوراه غير منشورة، جامعة اليرموك، اربد، الأردن.

الريان، عادل. (2010). دلالة الفروق في تحصيل طلبة الصف السادس الأساسي في الاختبار الوطني لمقرر الرياضيات وفقا لبعض المتغيرات. *مجلة جامعة الأقصى سلسلة العلوم الإنسانية المجلد الرابع* عشر، العدد الأول، ص 144-173.

الشريفين، نضال. (2009). بناء الاختبارات. ورقة مقدمة *في البرنامج التدريبي لأعضاء هيئة التدريس في جامعة أم القرى*، 7 تموز ،2009.

الصمادي، عزت. (2009). الاختبارات المحوسبة وبنوك الأسئلة. ورقة مقدمة *في البرنامج التدريبي لأعضاء هيئة التدريس في جامعة أم القرى*، 8 تموز، 2009.

علام، صلاح الدين محمود. (2006). *القياس والتقويم التربوي والنفسي أساسياته، وتطبيقاته، وتوجهاته المعاصرة* (ط2). القاهرة: دار الفكر العربي.

عوده، أحمد سليمان. (2010). *القياس والتقويم في العملية التدريسية* (ط4). إربد: دار الأمل.

عودة، أحمد، ملكاوي، فتحي. (1992). *أساسيات البحث العلمي*. إربد، الكتاني.

الفار، إبراهيم. (2000). *تربويات الحاسوب وتحديات مطلع القرن الحادي والعشرين* (ط.2).القاهرة، دار الفكر العربي.

مبارك، وائل. (2010). *"الأداء التفاضلي لفقرات اختبار العلوم فــي الدراســة الدوليــة بيــزا 2006"*. أطروحة دكتوراه غير منشورة، جامعة اليرموك، اربد، الأردن.

المومني، محمد ضيف الله. (2008). *"أثر الصورة التي تقدم فيها نصوص الاستماع باختبــار مهارة الاستماع باللغة الإنجليزية على الخصائص السيكومترية للاختبار وفقراته"*. أطروحة دكتوراه غير منشورة، جامعة اليرموك، اربد، الأردن.

المومني، صبحه. (2012). *" أثر طريقة تقديم اختبار اللغة الانجليزية للصف الخامس الأساسي على تقديرات القدرة وخصائص الاختبار وخصائص فقراته السيكومترية"*. رسالة ماجستير غير منشورة، جامعة اليرموك، اربد، الأردن.

وزارة التربية والتعليم. (2009).مشروع تطوير التعليم نحو اقتصاد المعرفة ERfKE II. استرجع في 20 تموز، 2013 من

http://www.moe.gov.jo/Projects/ProjectMenuDetails.aspx?MenuID=1&ProjectID=2

وزارة التربية والتعليم. (2011). *التقرير الإحصائي لنتائج الاختبار الوطني لضبط نوعية التعليم للصف العاشر الأساسي للعام الدراسي 2010-2011.*

وزارة التربية والتعليم. (2011). *تعليمات الاختبار الوطني لضبط نوعية التعليم للصف العاشر الأساسي للعام الدراسي 2010-2011.*

ثانيًا: المراجع الأجنبية:

Adarraga, P. & Zaccagnini, J.L.(1992). DAI: A knowledge-based system for diagnosing autism. A case study on the application of artificial intelligence to psychology. *European Journal of Psychological Assessment, 8*, 25 – 46.

Albanese, M. A., & Forsyth, R. A.(1984). The one-two and modified two- Parameter latent trait models: An empirical study of relative fit. *Education and Psychological Measurement, 44*, 229 – 246.

Allen, Mary J. and Yen, Wendy M. (1979). *Introduction to Measurement Theory.* California: Books/Cole Publishing Company.

Angoff, W. H. (1982). Use of difficulty and discrimination indices for detecting item bias. In R. A. Berk (Ed.), *Handbook of methods for detecting test bias* (pp. 96-116). Baltimore: The Johns Hopkins University Press.

AYDIN, S. (2006). The effect of computers on the test and inter-rater reliability of writing tests of ESL learners. *The Turkish Online Journal of Educational Technology – Tojet, ISSN: 1303-6521 volume 5* Issue 1 Article 9.

Baker, F. B. (1993). *EQUATE2: Computer program for equating two metrics in item response theory [Computer program]*. Madison: University of Wisconsin, Laboratory of Experimental Design.

Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., Yan, F. (2008). Does it Matter if I Take My Mathematics Test on Computer? A Second Empirical Study of Mode Effects in NAEP. *The Journal of Technology, Learning, and Assessment, 6* (9), 3 – 38.

Berk. R.A. (1982). Introduction. In BERk. R.A (Ed.), *Handbook of methods for detecting test bias* (pp. 1-9). Baltimore: The Johns Hopkins University Press.

Bodmann, Shawn M., Robinson, Daniel H.(2004). Speed and Performance Differences among Computer-Based and Paper-Pencil Tests. *Journal of Educational Computing Research, 1*(1), pp. 51- 60.

Boyd, A.(2003). *Strategies for Controlling Test let Exposure Rates in Computerized Adaptive Testing Systems*. Unpublished PhD Thesis University of Texas at Austin, Canterbury.

Bracey, G.(1990). *Computerized Testing: A Possible Alternative to Paper & Pencil?* Electronic Learning, February: 16-17.

Camili, G.& Shepared, L.(1994). *Methods for identifying bias test item*. Stage publication, USA.

Candell, G. L., & Drasgow, F.(1988). An iterative procedure for linking metrics and assessing item bias in item response theory. *Applied Psychological Measurement, 12*, 253–260.

Chung Wang, W. & Hui Su, y.(2004). Effects of average signed area between two item characteristic curves and purification procedures on the DIF detection via the Mantel–Haenszel method. *Applied Measurement in Education, 17* (2), 113 – 144.

Crocker, L.& Algina, J.(1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart & Winston.

Davidson, P.(2003). *The Equivalence of Paper-Based & Computer-Based Tests*. IATEFL Testing, Evaluation & Assessment SIG Newsletter

De Ayala, J., Kim, S., Stapletion & Dayton, C.(1999). A Re conceptualization of differential item functioning *paper presented at the Annual meeting of the American Education Research Association, Montreal, Canada*, April, 19-22.

De Gruijter, D.N.,& Vander Kamp, L.(2005). Statistical test theory for education and psychology. *Retrieved October 12, 2011, from http//irt.com.ne.kr/data/test theory.pdf*

Durance, N. & Holland, P.(1993). DIF detection and description Mantel–Haenszel standardization. *Educational testing service princetion, N.J. (QAT24225) ED287526*.

Ellis, B. & Raju, S.(2003). Test and item bias: what aren't and how to detect them measuring up. Assessment issues for teachers, counselors and administrator, (*Eric Document Reproduction Service No. ED 480042.*

Embretson, S.& Reise, S.(2000). *item response theory for psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Fraser, C. (1988). *NOHARM: an IPM PC computer program for fitting both unidimensional and multidimensional normal ogive models of latent triat theory.*Armidale, Australia, the university of new England

George, David .(2005). *Constructing Classroom Examination*. Commonwealth Publishers, New Delhi.

Gu, L.&. Drake ,S.& Wolfe, E.W.(2006). Differential Item Functioning of GRE Mathematics Items Across Computerized and Paper-and-Pencil Testing Media. *Journal of Technology, Learning, and Assessment, 5*, 4 - 30.

Hambleton, R. & Rogers, J.(1989). Detecting potentially bias test items comparison of IRT aria and Mantel–Haenszel methods. *Applied Measurement in Education, 2* (4), 313 –334.

Hambleton, R.& Rogers.(1995). Item bias review practical assessment, research and evaluation. *Retrieved July, 31, 2013 from http//edresearch.org/pare/getvn.ASP=4&n=6.*

Hambleton, R.& Swaminathan, H.& Rogers, J.(1991). *Fundamentals of item response theory.* Newbury Perk California: Stage publications.

Hambleton, R.& Swaminathan, H.(1985). *Item response theory: principles and applications.* Boston, MA: Kluwer- Nijhof.

Hambleton, R., Cla., Mazor, K. & Jones, R. (1993). Advanced in the detecting of differential functioning test item. (*Eric Document Reproduction Service No. ED 356264*).

Han, K. T. (2007). *IRTEQ: Windows application that implements IRT scaling and equating [computer program]*. Amherst, MA: University of Massachusetts Amherst, Center for Educational Assessment. Available for download at http://www.umass.edu/remp/software/irteq/.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied psychological Measurement, 9* (2), 139 - 164.

Higgins, Jennifer, Patterson, Margaret, Bozman, Martha & Katz, Michael.(2010). Examining the Feasibility and Effect of Transitioning GED Tests to Computer. *The Journal of Technology, Learning, and Assessment, 10* (2), 1 - 33.

Horkay, N., Bennett, R.E, Allen, N., Kaplan, B., &,Yan, F.(2006). Does it Matter if I Take My Writing Test on Computer? An Empirical Study of Mode Effects in NAEP. *The Journal of Technology, Learning, and Assessment, 5* (2), 1 - 49.

Houser, Carl, Kingsbury, Gage.(2004). Differential Item Functioning and Differential Test Functioning In the Idaho Standards Achievement Tests for Spring 2003. *Internship in Measurement and Statistics (EDMS 889).*

Hulin, C., Drasgow, F. & Person, C.(1983). *Item response theory: Application to psychological Measurement*. Homewood, IL: Dow and Jones Irwin.

Ironson, G. (1982). Using Item Response Theory to measure bias. In R.K. Hambeleton (Ed.), *Application of Item Response Theory* (pp. 155-175). Vancouver: Educational Research Institute of British Columbia

Jasper, F.(2010). Applied dimensionality and test structure assessment with–the START–M mathematics test. *Journal of educational and Psychological assessment, 6* (1), 109 –126.

Jennifer L., Kobrin & John W. Young.(2003). The Cognitive Equivalence of reading Comprehension Test Items Via Computerized and Paper-and-Pencil Administration. *Applied Measurement in Education, 16* (2), 115 –140.

Jensen, R.(1980). *Bias in mental testing*. New York: Macmillan Publishing CO. Inc.

Jodoin, M.G. (2003). Measurement efficiency of innovative item formats in computer-based testing. *Journal of Educational Measurement, 40* (1), 1– 15.

Johnson, Martin, Green, Sylvia.(2006). On-Line Mathematics Assessment: The Impact of Mode on Performance and Question Answering Strategies. *The Journal of Technology, Learning, and Assessment, 4* (5), 2 – 35.

Karakaya, İsmail. (2012). An Investigation of Item Bias in Science & Technology Subtests and Mathematic Subtests in Level Determination Exam (LDE). *Educational Sciences: Theory & Practice, 12* (1) • Winter • 222-229 ©2012 Educational Consultancy and Research Center www.edam.com.tr/estp.

Kim, Do-Hong & Huynh, Huynh.(2010). Equivalence of Paper-and-Pencil and Online Administration Modes of the Statewide English Test for Students With and Without Disabilities. *Educational Assessment, 15*, 107 – 121.

Kim, S., Cohen, A.& Lin, Y.(2005). *LDID: A Computer program for local dependence indices for dichotomous items*. Version 1.0.

Kobrin, J. L.(2000). An Investigation of the Cognitive Equivalence of Computerized & Paper-Pencil Reading Comprehension Test Items. *Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans*, Louisiana. (ED 442 836).

Kolen, M. J., & Brennan, R. L.(2004). *Test equating, scaling, and linking: Methods and practices (2nd ed.).* New York: Springer.

Lee, K., & Oshima, T. C. (1996). IPLINK: Multidimensional and unidimensional item parameter linking in item response theory. *Applied Psychological Measurement, 20*, 230. Available for download at http://education.gsu.edu/eps/4493.html

Lord, F.m. & Nonvick, M.R.(1968). *Statistical theories of mental test scores*. Reading, MA: Addison – Wesley.

Mc Donald, R.(1982). Linear versus non–linear models in item response theory . *Applied psychological measurement, 6*, 379 - 396.

Meara, K., Robin, F.& Sireci, S.(2005). Using Multidimensional Scaling to Asses the Dimensionality of Dichotomous Item Data. *Multivariate Behavioral Research, 35* (2), 229 - 259.

Oshima, T. C., Raju, N. S., & Nanda (2006). A new method for assessing the statistical significance in the differential functioning of items and tests (DFIT) framework. *Journal of Educational Measurement, 43*, 1-17.

Osterlind, s. (1983). *Test item bias*. Beverly Hills: Sage publications.

Pae, T.(2004). DIF for examinees with different academic background. *Language testing, 21*(1), 53 -72.

Perrone, M. (2006). Differential item functioning and item bias: Critical Consideration In TESOL *and Applied Linguistics, 6* (2), the form retrieved, May, 2013 from, http://www.tc.collumbia.edu/tesolalwebjournal.

Piaw, Chua Yan. (2010). *Comparisons between computer-based testing and paper-pencil testing: testing effect, test scores, testing time and testing motivation*. Institute of Educational Studies, University of Malaya, 50603 Kuala Lumpur, Malaysia.

Pommerich, Mary; Burden, Timothy. (2000). From Simulation to Application: Examinees React to Computerized Testing. *Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans*, LA.

Pomplun, M., Frey, S & Becker, D. F.(2002). The Score Equivalence of Paper-&-Pencil & Computerized Versions of a Speeded Test of Reading Comprehension. *Educational & Psychological Measurement, 62* (2): 337- 354.

Puhan, G.& Boughton, K.& Kim ,S.(2007). Examining Differences in Examinee Performance in Paper and Pencil and Computerized Testing. *Journal of Technology, Learning, and Assessment, 6*, 5 - 20.

Raju, N.(1988). The area between two item characteristic curves. *psychometrika, 53* (4), 495 - 502.

Raju, N. S., Oshima, T. C., & Wolach, A. (2009). *Differential functioning of items and tests (DFIT): Dichotomous and polytomous [Computer program]*. Chicago: Illinois Institute of Technology.

Raju, N. S., Oshima, T. C., Fortmann, K., Nering, M., & Kim, W. (2006). The new significance test for Raju's polytomous DFIT. *Paper presented at the New Directions in Psychological Measurement with Model-Based Approaches at Georgia Institute of Technology in Atlanta*, GA.

Raju, N. S., van Der Linden, W. J., & Fleer, P. F. (1995). An IRT- based Internal measure of test bias with applications for differential item functioning. *Applied Psychological Measurement, 19*, 353 – 368.

Roever, C. (2005). "That's not fair" Fairness, bias, and differential item functioning in language testing. *Retrieved, February, 6, 2011, from http://www.hawaii.edu/2ve/brounbag.doc.*

Rudner, L. M., Getson, P. R., & Knight, D. L. (1980a). Biased item detection techniques. *Journal of Educational Statistics, 5*, 213-233.

Rudner, L. M., Getson, P. R., & Knight, D. L. (1980b). A Monte Carlo comparison of seven biased item detection techniques. *Journal of Educational Measurement, 17*, 1-11.

Sandifer, M.(2001).Testing, Testing Avoiding bias in testing. *Retrieved June 15, 2010 ,from the national conference of Bar examiners: http://www.ncbex.org/uploads/userdocrepose/700401-testing.pdf.*

Schwarz, Richard D., Rich, Changhua, Podrabsky, Tracy.(2003). A DIF Analysis of Item-Level Mode Effects for Computerized and Paper-and-Pencil Tests. *Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago,* IL, April 22-24.

Shen, L. (1997). Quantifying item dependency by Fishers Z.*(Eric Document Reproduction Service No. ED 4102. 41)*.

Sheng, Y. (2005). *Bayesian analysis of hierarchical IRT models: comparing and combining the Unidimensional & Multi-Unidimensional IRT models*. Un published doctoral, University of Missouri, Columbia.

Stoneberg Jr, Bert D. (2004). A Study of Gender-Based and Ethnic-Based Differential Item Functioning (DIF) in the Spring 2003 Idaho Standards Achievement Tests Applying the Simultaneous Bias Test (SIBTEST) and the Mantel-Haenszel Chi Square Test. *Internship in Measurement and Statistics (EDMS 889)*

Swaminathan, H. & Rogers, J.H. (1990). Detecting differential item functioning using Logistic regression procedures. *Journal of Educational Measurement. 27* (4), 361 – 370.

Thompson, S., Thurlow, M., & Moore, M. (2002). *Using computer based tests with students with disabilities (Policy Directions No. 15*). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Viser, M. (1998). Adaptive testing comparisons. South African university. *Journal of Psychology, 63*, vol, 28.

Webster, J. & Compeau, D.(1996). Computer-Assisted versus Paper-Pencil Administration of Questionnaires. Behavior Research Methods, Instruments, & Computers, 28(4): 567-576

Westmeyer, H. & HagebÖck, J.(1992). Computer-Assisted Assessment: A Normative Perspective. *European Journal of Psychological Assessment, 8*, 1-17.

Wise, S., Plake, B., Johnson, P. & Roos, (1992). A comparison of Self-Adapted and Computerized Adaptive Tests, *Journal of Education Measurement, 29*,4,329 - 339.

Wise, S.L.,& Plake, B.S.(1990). Computer-based testing in higher education. *Measurement and Evaluation in Counseling and Development, 23*(1), 3 -10.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D.(2002). *BILOG-MG3 [Computer software].* St.Paul, MN: Assessment Systems Corporation.

Zimowski, Michele F., Muraki, Eiji, Mislevy, R., Robert, J.& Back, D.(1996). BILOG-MG 3: *Multiple group IRT analysis and test maintenance for binary items* (Computer program). Chicago: Scientific software. 17, No. 1

Zumbo, B, D.(1999). *A handbook on the theory and methods of differential item functioning.* Ottawa: Directorate of Human resources research and evaluation, Department of National Defense.

Zumbo, B, D. (2007). Three generations of DIF analysis: considering where it has been, where it is now, and where it is going. *Language Assessment quarterly , 4*(2). 223 – 233.

الملاحق

ملحق (1)



إدارة الامتحانات والاختبارات
مديرية الاختبارات
قسم القياس والتقويم

الاختبار الوطني لضبط نوعية التعليم

الصف العاشر

للعام الدراسي ٢٠١١

الورقة الثانية

(اللغة الإنجليزية والعلوم)

رمز الورقة (١٢)

اقرأ هذه التعليمات ولا تفتح كراسة الاختبار حتى يسمح لك مشرف القاعة بالبدء.

تعليمات الاختبار:

- يتكون هذا الاختبار من ٦٠ فقرة من نوع الاختيار من متعدد لمبحثي اللغة الإنجليزية والعلوم.
- لكل فقرة أربعة بدائل واحد منها فقط صحيح.
- ظلل رمز الإجابة الصحيحة في نموذج الإجابة المرفق مستخدما قلم الرصاص.
- مدة الاختبار ساعة ونصف.
- إذا أردت تغيير الإجابة عن فقرة ما استخدم الممحاة.
- أجب عن جميع الأسئلة بعناية ودقة .
- اسأل مشرف القاعة إذا شعرت أنك بحاجة إلى مساعدة لفهم سؤال ما.

ملاحظة: اتبع الإرشادات المكتوبة على نموذج الإجابة.

Read the following text, and then answer the questions below:

Teenagers play computer games all over the world in order to satisfy their needs. Firstly, they play for entertainment. Secondly, they play them to sharpen their computer skills. Finally, they play them to use their higher mental abilities to solve problems.

Unlike other games, computer games players perform them individually; their partner near the computer set itself only. They hurt their eyes and their backbones by sitting for a long time in front of the set.

Select the choice that best completes each of the following sentences according to the text above:

١. Computer games have ________________ .

a- entertainment advantages. b- unsatisfactory advantages.
c- mental disadvantages. d- mental skills disadvantages.

٢. Computer games players can play them ______________ .

a- alone b- in pairs. c- in a team d- in a group of people

Read the following text, and then answer the questions below:

It is difficult to imagine life without computers. They are a source of information, education and entertainment, but hackers spend their time playing with computer data in all parts of cyberspace. They know how to trick people using their computers.

Most hackers still think that their activities are a game. They often meet festivals to take part in discussions, share their experiences, meet other hackers and generally have a good time. These meetings are organized in well-known places like Las Vegas or Berlin. However, what the hackers do at such festivals is a secret and often many of their activities take place at night.

Recently, hacking has started to increase. Hackers are getting in to computers systems and stealing or destroying information.

Select the choice that best completes each of the following sentences according to the text above:

٣. Hackers give computer users ______________ .

a- entertainment b- stress and problems
c- useful facts d- interesting discussion

٤. In future years, the number of hackers ____________ .

a- will not change b- will be bigger
c- will decrease d- will not be a problem

٥. Hackers spend their time________________ .

a- making computer games b- destroying computer viruses
c- changing computer data d- making websites safer

٦. The best title for the text above is______________ .

a- Computers b- Internet Safety
c- Hackers d- Online Communication

Read the table, and then complete the sentences with the correct choice:

Data \ City	London	Mexico City
Location	On the river Thames in the south of England	On a plain in the center of Mexico
Population	٧.٥ million	١٨ million
Language	٣٠٠ languages including English	Spanish
Sights	Buckingham Palace	Ancient Aztec ruins

٧. Mexico is bigger than London because its population is __________ .

a- ٧,٥ million b- ١,٨ million c- ٣٠٠ million d- ١٨ million

١

٨. **Both cities can be tourist attractions because they have______________.**

a-rivers b- locations c- sights d- languages

٩. **Bereavement means ________________**

a- feasting b- thick fermented liquor

c- cook in hot oil d- loss by death.

١٠. **My specialty is bread with falafel which can be filled up with __________ and vegetables.**

a- cakes b- pickles c- rice d- soup

١١. **Select the choice that best completes each of the following sentences:**

> **If you believe that all things happen for the best, you are most likely _________________person.**

a- an optimist b- a stressful c- a virtual d- an arrogant

١٢. **The Sydney Opera House is less _______________than Burj Al-Arab.**

a- sophisticated b- sophistication c-sophisticating d-sophisticate

Read the dictionary extract, and then complete the sentence with the correct choice:

> **Compete** (v) *take part in a contest*
> **Competition** (n) *rivalry, opposition*
> **Competitive** (adj) *demanding, challenging*

١٣. **Education is becoming more developed and_______________.**

a- competition b- competitive c- compete d- competitions.

١٤. **Select the choice that best completes the following sentence.**

> **A : How ____________________________ .**
> **B : Two hours.**

a- long new is the game? b- new game the is long?

c- long is the new game? d- the new game is long?

١٥. **Select the choice that best completes the following sentence.**

> **A : ________________________________?**
> **B : Yes, I went shopping with my classmates in Irbid yesterday.**

a- Was you go shopping with your classmates in Irbid yesterday?

b- Did you go shopping with your classmates in Irbid yesterday?

c- Did you went shopping with your classmates in Irbid yesterday?

d- Were you go shopping with your classmates in Irbid yesterday?

١٦. **You asked your friend to bring you " dish dash " from Mecca.**

> **He said: please, forgive me; I did not bring you the " dish dash" .**

What would you say to him?

a- Never mind. b- You must bring it.

c- It is too late to apologize. d- That's great! Well done.

١٧. **Select the appropriate choice to complete the sentence below:**

Ali missed the bus. He is ٢٠ minutes late.
He is very sorry because the teacher says: ______________________.

a- You must come early . b- Shall you come early ?
c- Have you come early ? d- You must not come early .

١٨. **Select the choice that best completes the following sentence.**

Ibn-Battuta left his home ____________ Mecca at the age of twenty.

a- through b- from c- for d- of

١٩. **Select the choice that best completes the space in the following dialogue :**

Ahmad : Excuse me
Sami : Yes?
Ahmad : ______________________________.
Sami : Yes, of course. Go along the road as far as the traffic lights. Then turn left into the Height Street and it is on the right.

a. What about going to the bank, please?
b. Could you tell me how to get to the bank, please?
c. Would you like to come out with me to the bank, please?
d. Will you give me some information a bout the new bank, please?

Read the following paragraph carefully, and then answer the question that follows:

Ali was driving fast. He didn't notice there were traffic comes on the street. He hit them and made a terrible accident.

٢٠. **Which one of the following sentences would you say to Ali?**
a- You shouldn't have driven fast. b- You mightn't have driven fast.
c- You wouldn't have driven fast. d- You can't have driven fast.

٢١. **Select the choice that best completes the following sentence.**

Ibrahim is ______________ because he has bad luck .

a- lucky b- unlucky c- uncle d- unlock

٢٢. **Choose the correctly punctuated sentence.**
a- Zeidan is a brilliant, fast skillful and intelligent footballer.
b- Zeidan is a brilliant fast, skillful, and intelligent footballer.
c- Zeidan is a brilliant fast skillful and intelligent footballer.
d- Zeidan is a brilliant, fast, skillful and intelligent footballer.

٢٣. **Choose the correctly punctuated sentence.**
a- Its a beautiful day isn't it?
b- It's a beautiful day. Isn't it?
c- It's a beautiful day, isn't it?
d- It's a beautiful day Isn't, it?

٣

٢٤. **Complete the following sentence with the correctly spelt word.**

> **Serious games are ____________ to convey information of a learning experience of some sort to the player.**

a- desined b- designed c- desinged d- disinged

٢٥. **Which word in the sentence below should be capitalized?**

> **At last, and at the age of twenty he went to paris to become a lawyer.**

a- Age b- Twenty c- Paris d- Lawyer

٢٦. **Which word in the sentence below should be capitalized?**

> **Sami was one of the police officers at the Park Zoo last september when the animals' cage fell down.**

a- Police b- Officers c- September d- Cage

٢٧. **Nour has read thirty books ____________ ٢٠٠٥.**

a- since b- for c- already d- yet

٢٨. **Which of the following is the correct translation of the Arabic sentence below?**

> **هل تحب أن تكون بطلا رياضيا عالميا؟**

a- When you like to be a world champion in a sport?
b- Would you like to be a world champion in a sport?
c- Do you like to be a world champion in a sport?
d- Will you like to be a world champion in a sport?

٢٩. **Which of the following is the correct translation of the English sentence below?**

> **Laptop computers will get smaller and become more powerful.**

a- الحواسيب المحمولة اصغر حجما وأكثر طاقة
b- الحواسيب المحمولة سوف تصبح اكبر حجما وأكثر طاقة
c- الحواسيب المحمولة سوف تصبح اصغر حجما و اقل طاقة
d- الحواسيب المحمولة سوف تصبح اصغر حجما وأكثر طاقة

٣٠. **Which of the following is the correct translation of the English sentence below ?**

> **Television is one of the greatest inventions in the twentieth century.**

a- التلفاز هو إحدى أعظم الاختراعات في القرن العشرين .
b- التلفاز أعظم اختراع في القرن العشرين .
c- اختراع التلفاز عظيم في القرن العشرين .
d- إحدى أعظم الاختراعات في القرن العشرين هو التلفاز .

انتهت أسئلة اللغة الإنجليزية

تابع أسئلة العلوم من (٣١) الى (٦٠)

٣١- قام فني مختبر بتحديد عدد خلايا البكتيريا الناتجة من تكاثر ١٠٠٠ خلية من نوع يدعى *Escherichia coli* بالانشطار الثنائي، فوجد أن البكتيريا تتكاثر كما في الرسم البياني الآتي. اعتماداً على الرسم كم يصبح عدد خلايا البكتيريا بعد مضي ساعة ونصف؟

عدد الخلايا البكتيرية: ٢٠٠٠، ٤٠٠٠، ٦٠٠٠، ٨٠٠٠
الزمن (دقيقة): ٣٠، ٦٠، ٩٠

أ) ٤٠٠٠
ب) ٨٠٠٠
ج) ١٢٠٠٠
د) ١٦٠٠٠

٣٢- تعيش الحية ذات الطوق والحية السامة في المنطقة نفسها، غير أن الأولى تعيش في الأماكن الرطبة، أما الثانية فتعيش في الأماكن الأقل رطوبة، وبالرغم من أنهما غير منعزلين جغرافياً وينتميان لنفس الجنس (من نوعين مختلفين) إلا أنهما نادراً ما يلتقيان. ما نوع الانعزال التكاثري في هذه الحالة؟

أ) فصلي　　ب) سلوكي　　ج) تركيبي　　د) بيئي

٣٣- تحتوي خلايا الكائن الحي الجسمية على عدد ثنائي من الكروموسومات، تترتب بصورة أزواج في أحد أطوار دورة حياة الخلية. أي الأشكال الآتية يمثل الترتيب الصحيح للجينات في زوج الكروموسومات المتقابلة؟

أ) X Y Z / X Y Z
ب) X Y Z / Z X Y
ج) Y Z X / X Y Z
د) Y X Z / X Y Z

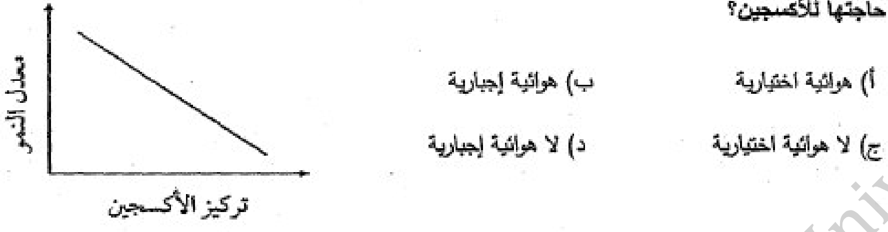
٣٤- لاستقصاء أثر الغذاء في نمو أحد أنواع البكتيريا، تم زراعة عينة من البكتيريا في وسط غذائي مناسب، وأخذت قراءات في فترات زمنية محددة لأعداد الخلايا البكتيرية الموجودة في الوسط الغذائي، والشكل الآتي يبين النتائج التي تم التوصل إليها. أي الأرقام في الشكل يشير إلى المرحلة التي كان فيها أعلى معدل لنمو البكتيريا؟

معدل النمو
١ ٢ ٣ ٤
الزمن

أ) ١　　ب) ٢
ج) ٣　　د) ٤

٥

٣٥- أجرى طالب تجربة لاستقصاء أثر تركيز الأكسجين على نمو نوع من البكتيريا يدعى ***Clostridium botulinum*** التي تسبب التسمم الغذائي، ومثّل النتائج بيانياً كما في الشكل الآتي. ما نوع هذه البكتيريا حسب حاجتها للأكسجين؟



أ) هوائية اختيارية　　ب) هوائية إجبارية

ج) لا هوائية اختيارية　　د) لا هوائية إجبارية

٣٦- أراد شخص أن يتأكد من صفة ما بأنها سائدة، واستطاع من خلال دراسته لانتشار هذه الصفة بين أفراد المجتمع استخراج نسبة ظهورها بين أفراد ذلك المجتمع. أي النسب الآتية تؤكد أن الصفة المدروسة سائدة؟

أ) $\frac{١}{٨}$　　ب) $\frac{٣}{٨}$　　ج) $\frac{٥}{٨}$　　د) $\frac{٦}{٨}$

٣٧- قام مزارع برش حظيرة أبقاره بمحلول مبيد حشري للقضاء على الذباب المنتشر فيها، فلاحظ عند الاستخدام الأول أن ٩٠% من الذباب تقريباً قد قتلت، وبعد شهرين استخدم نفس المبيد الحشري فلاحظ أن ٦٥% من الذباب تقريباً قد قتلت. ما تفسير التناقص في فعالية المبيد الحشري؟

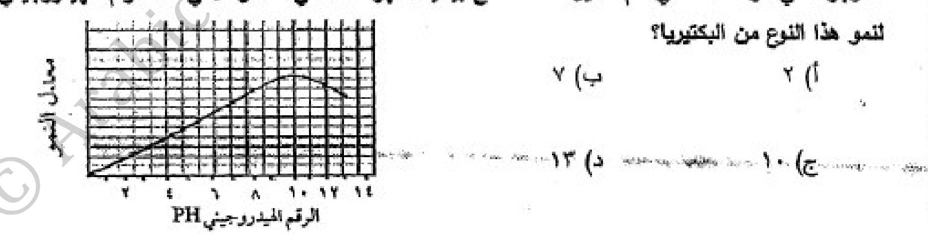
أ) أثّر المبيد الحشري في الذباب البالغ بينما كان تأثيره أقل في الذباب غير البالغ.

ب) تعلّم بعض الذباب كيف يقاوم المبيد الحشري، وعلّم باقي الذباب.

ج) تكاثر الذباب الذي قاوم المبيد عند الاستخدام الأول، وورّث الصفة إلى الأبناء.

د) أحدث المبيد الحشري طفرة في القناة الهضمية لبعض الذباب مما جعلها تفككه.

٣٨- لاستقصاء أثر الرقم الهيدروجيني PH في نمو أحد أنواع البكتيريا، تم زراعة عينة من البكتيريا في وسط غذائي وتم تعديل الرقم الهيدروجيني للوسط الغذائي خلال فترات زمنية محددة، وأخذت قراءات لمعدل نمو الخلايا البكتيرية الموجودة في الوسط الغذائي، تم تمثيل هذه النتائج بيانياً فظهرت كما في الشكل الآتي. ما الرقم الهيدروجيني الأمثل لنمو هذا النوع من البكتيريا؟



أ) ٢　　ب) ٧

ج) ١٠　　د) ١٣

٣٩- عدسة محدبة قوتها(٤ ديوبتر)، إذا وضع جسم على مسافة (٥٠ سم) منها، على أي مسافة يتكون خيال للجسم؟

أ) ٢٥ سم　　ب) ٣٠ سم　　ج) ٤٠ سم　　د) ٥٠ سم

٤٠- تبين الأشكال الآتية شحنات كهربائية، والرموز (ف) و(٢ ف) تشير إلى المسافة و ضعفها على التوالي. في أي الأشكال الآتية تكون قوى التجاذب بين الشحنتين أكبر ما يمكن؟

أ) +١ ـــ ف ـــ -١

ب) +٢ ـــ ف ـــ -٢

ج) +١ ـــ ٢ ف ـــ -١

د) +٢ ـــ ٢ ف ـــ -٢

٤١- وضع جسم على شكل سهم أمام مرآة مقعرة بعدها البؤري (ب) ومركز تكورها (م)، أي الأشكال الآتية يبين الموضع الصحيح لخيال رأس السهم؟

٤٢- في الشكل الآتي كم تبلغ مقاومة الموصل؟

أ) ١٠ أوم ب) ٤ أوم

ج) ٥ أوم د) ٢,٥ أوم

٤٣- إذا كان التيار المار في المصباح (٢ كولوم/ ثانية) يستهلك طاقة مقدارها (١٠جول). ما قيمة الجهد الكهربائي؟

أ) ٢ فولت ب) ٥ فولت ج) ١٠ فولت د) ٢٠ فولت

٤٤- أي الدارات الكهربائية الآتية يمكنك استخدامها لقياس التيار الكهربائي وفرق الجهد عبر المقاومة ٣Ω ؟

٧

٤٥- أجرى طالب في مختبر الفيزياء تجربة لتوليد تيار حثي باستخدام ملف تم تحريكه بين قطبي مغناطيس لكنه لاحظ عدم تولد تيار حثي. ما تفسير ذلك؟

أ) مستوى الملف عمودياً على خطوط المجال. ب) تغير عدد خطوط المجال المغناطيسي.

ج) حركة الملف والمجال المغناطيسي بنفس الاتجاه. د) اختراق الملف لخطوط المجال المغناطيسي.

٤٦- وقف شخص أمام مرآتين مستويتين بينهما زاوية مقدارها (٦٠ درجة)، ما عدد الأخيلة المتكونة له؟

أ) ١٥ ب) ١٢ ج) ٥ د) ٣

٤٧- أي الأشكال الآتية يحدث فيها انكسار كلي داخلي للضوء ؟

أ) ن زجاج = ١,٥ ، ن هواء = ١ (θ_c)

ب) ن هواء = ١ ، ن زجاج = ١,٥ (θ_c)

ج) ن ماء = ١,٣٣ ، ن زجاج = ١,٥ (θ_c)

د) ن هواء = ١ ، ن ماء = ١,٣٣ (θ_c)

٤٨- إذا علمت أن أحد مصانع الأمونيا ينتج يومياً (١٧٠ كغم)، وفق المعادلة الموزونة الآتية، ما كتلة النيتروجين بالكيلوغرام اللازمة لإنتاج الأمونيا ($NH_٣$) يومياً؟ الكتلة الذرية (N = ١٤ ، H = ١)

$$٣H_٢ + N_٢ \longrightarrow ٢NH_٣$$

أ) ١٤ ب) ١٤٠ ج) ٧ د) ٧٠

٤٩- يستهلك لاعب كرة سلة طاقة قدرها (٢١٠٠ كيلو جول/ ساعة تدريب)، إذا علمت أن عملية احتراق سكر الغلوكوز في جسم اللاعب لتزويده بالطاقة تتم وفق المعادلة الموزونة الآتية :(الكتلة المولية للسكر= ١٨٠غم/ مول).

$$C_٦H_{١٢}O_٦ + ٦O_٢ \longrightarrow ٦CO_٢ + ٦H_٢O + ٢٨٠٠ \ KJ$$

ما أقل كمية من السكر يحتاجها اللاعب للتدرب لمدة ساعتين؟

أ) ٢٧٠ غم ب) ١٣٥ غم ج) ١٨٠ غم د) ٣٦٠ غم

٥٠- أي الآتية تترابط ذراته بروابط فلزية؟

الأعداد الذرية (Na = ١١ ، Cl = ١٧ ، Mg = ١٢ ، C = ٦ ، H = ١ ، O = ٨)

أ) NaCl ب) $CH_٤$ ج) Mg د) $CO_٢$

٥١- مستعيناً بالجزء الآتي من الجدول الدوري، ما رمز العنصر الذي يقع في الدورة الثالثة والمجموعة الرابعة؟

العدد الذري

الكتلة الذرية

أ) C ب) Si ج) Al د) B

٥٢- تمكن أحد الباحثين من تحديد الصيغة الأولية لمركب ما بـ ($CH_٢$). إذا علمت أن الكتلة المولية لهذا المركب تساوي ٥٦ غم / مول ، ما الصيغة الجزيئية للمركب؟ (الكتل الذرية : C =١٢ ، H =١)

٨

أ) C_2H_4 ب) C_3H_6 ج) C_4H_8 د) C_5H_{10}

٥٣- في مختبر الكيمياء ، قام أحد الطلبه بحرق شريط من المغنيسيوم حيث لاحظ تكون مسحوق أبيض اللون، إذا علمت أن المركب الناتج هو أكسيد المغنيسيوم، ما صيغته الكيميائية؟ (الأعداد الذرية: Mg = ١٢، O = ٨)

أ) Mg_2O ب) MgO_2 ج) Mg_2O_2 د) MgO

٥٤- رسم أربعة طلاب النماذج الآتية على السبورة لجزيء الماء (H_2O). أيها يمثل نموذج لويس الصحيح؟ (الأعداد الذرية : O = ٨ ، H = ١) .

أ) H — O — H   ب) H — Ö — H

ج) H — Ȯ — H   د) H —••O••— H

٥٥- يمثل الجدول الآتي بعض الخصائص الفيزيائية لأربع مواد. أي المواد (D ، C ، B ، A) تمثل مادة أيونية؟

المادة	درجة انصهارها (س°)	ذائبيتها في الماء	توصيلها للتيار الكهربائي في الحالة الصلبة
A	١٥٠	تذوب	رديء
B	٩٥٠	تذوب	رديء
C	١٥٠٠	لا تذوب	جيد
D	١٦٠	تذوب	رديء

أ) A   ب) B   ج) C   د) D

٥٦- يمثل الجدول الآتي المتوسط السنوي للضغط الجوي في عدد من محطات الرصد الجوي في الأردن، ما مقدار الضغط الجوي في محطة الرصد رقم (٤)؟

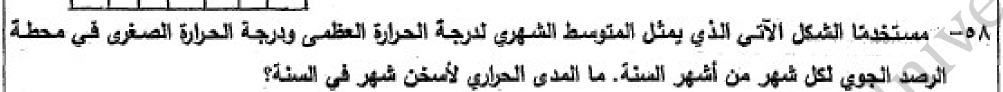
رقم المحطة	الارتفاع عن سطح البحر (متر)	المتوسط السنوي للضغط الجوي (مليبار)
١	٦١٦	٩٥٠
٢	٧٨٠	٩٢٥
٣	٦٧٢	٩٣٥
٤	٨٥٠	....

أ) ١٠٥٠ مليبار
ب) ١٠٢٠ مليبار
ج) ٩٨٠ مليبار
د) ٩١٠ مليبار

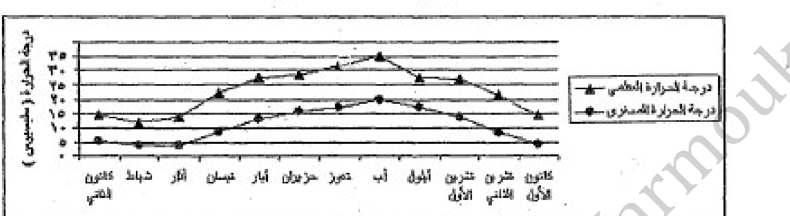
٥٧- يوضح الشكل الآتي مجال درجة الحرارة في منطقة ما في أحد الأيام، والخطوط المنحنية تمثل خطوط تساوي درجة الحرارة فوق تلك المنطقة، ما مقدار درجة الحرارة في المنطقة (و)؟

أ) ١١°س ب) ١٣°س

ج) ١٢°س د) ١٥°س



٥٨- مستخدمًا الشكل الآتي الذي يمثل المتوسط الشهري لدرجة الحرارة العظمى ودرجة الحرارة الصغرى في محطة الرصد الجوي لكل شهر من أشهر السنة. ما المدى الحراري لأسخن شهر في السنة؟



أ) ٥ س°

ب) ١٥ س°

ج) ٢٠ س°

د) ٣٥ س°

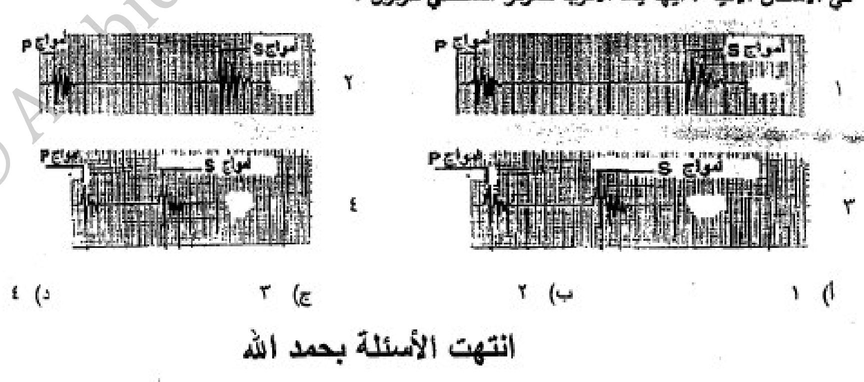
٥٩- لاحظ أحد علماء البراكين في جزيرة هاواي أن بعض البراكين ينساب منها الصهير بشكل هادئ، وبعضها الآخر يقذف اللابة والمواد البركانية في الجو بصورة انفجارات عنيفة. مستعيناً بالجدول الآتي أي البراكين المشار إليها بالأرقام (١، ٢، ٣، ٤) ينتج انفجارات عنيفة؟

البركان	نوع البركان	درجة الحرارة
١	قاعدية	١٢٠٠س°
٢	حمضية	٩٠٠ س°
٣	حمضية	٨٠٠س°
٤	قاعدية	١٠٠٠ س°

أ) ١ ب) ٢

ج) ٣ د) ٤

٦٠- حلل أحد المختصين في رصد الزلازل أربع مخططات زلزالية تم الحصول عليها من عدة محطات رصد زلزالي كما في الأشكال الآتية . أيها يعد الأقرب للمركز السطحي للزلزال ؟



أ) ١ ب) ٢ ج) ٣ د) ٤

انتهت الأسئلة بحمد الله

ملحق (2)



إدارة الامتحانات والاختبارات

مديرية الاختبارات

الاختبار الوطني لضبط نوعية التعليم

الصف العاشر

للعام الدراسي ٢٠١١

الورقة الأولى

(اللغة العربية والرياضيات)

رمز الورقة: (١١)

اقرأ هذه التعليمات ولا تفتح كراسة الاختبار حتى يسمح لك مشرف القاعة بالبدء.

تعليمات الاختبار:

- يتكون هذا الاختبار من ٦٠ فقرة من نوع الاختيار من متعدد لمبحثي اللغة العربية والرياضيات.
- لكل فقرة أربعة بدائل واحد فقط منها صحيح.
- ظلل رمز الإجابة الصحيحة في نموذج الإجابة المرفق مستخدما قلم الرصاص.
- مدة الاختبار ساعة ونصف.
- إذا أردت تغيير الإجابة عن فقرة ما فاستخدم الممحاة.
- أجب عن جميع الأسئلة بعناية ودقة .
- اسأل مشرف القاعة إذا شعرت أنك بحاجة إلى مساعدة لفهم سؤال ما.

ملاحظة: اتبع الإرشادات المكتوبة على نموذج الإجابة.

اقرأ الأبيات الآتية، ثم أجب عن الأسئلة (١-٥):

زَعموا بِأنَّ الصَّقْرَ صادَفَ مَرَّةً ... عُصفورَ بَرٍّ ساقَهُ التَّقْديرُ

فَتَكَلَّمَ العُصفورُ تَحْتَ جَناحِهِ ... والصَّقْرُ مُنْقَضٌّ عَلَيْهِ يَطيرُ

إنّي لِمِثْلِكَ لا أُتَمِّمُ لُقْمَةً ... وَلَئِنْ شُوِيتُ فَإنَّني لَحَقيرُ

فَتَهاوَنَ الصَّقْرُ المُدِلُّ بِصَيْدِهِ ... كَرَماً وَأَفْلَتَ ذَلِكَ العُصفورُ

١- تدور الأبيات حول موضوع واحد هو:

أ) شكوى الزمان ب) العفو عند المقدرة ج) الندم د)الاغتراب

٢- الأسلوب الذي استخدمه الشاعر في الأبيات السابقة هو:

أ) القصصيّ ب) الخطابيّ ج) التهكميّ د) المسرحيّ

٣- يعود الضمير (الهاء) في كلمة (بصيده) على:

أ) الصقر ب) جناح ج) العصفور د) لقمة

٤- وصف الشاعر الصقر بصفة هي:

أ) الشجاعة ب) الأمانة ج) الطيش د) الكرم

٥- شبّه الشاعر العصفور في البيت الثاني من الأبيات السابقة بـ:

أ) الصقر ب) الإنسان ج) الصياد د) النسر

٦- إحدى الآتية ليست من سمات المثل:

أ) الإيجاز ب) سهولة الحفظ ج) سهولة النقل د) السخرية

٧- يُضرب مثل (مَواعيدُ عُرْقوب) للدلالة على الإنسان:

أ) المراوغ ب) الحليم ج) اللئيم د) الصبور

٨- وَكَيْفَ أَرْمي لِساناً طالَما سُقِيَتْ ... بِشَهْدِهِ الْبيدُ مِنْ شَرْبٍ وَسُمّارِ؟

يتحدّث البيت السابق عن صفة في عنترة هي:

أ) شجاعته في البيداء ب) عفّته عن الشرب ج) مخالطته السمار د) عذوبة شعره

٩- عنصر مما يأتي ليس من عناصر القصة القصيرة:

أ) هدف المؤلّف ب) الشخوص ج) الحبكة الفنيّة د) المكان

١٠- الترتيب الصحيح للكلمات الآتية حسب ورودها في المعجم الوسيط هو:

أ) القفرة، الأيكة، ابتعد، انتصر ب) الأيكة، القفرة، انتصر، ابتعد

ج) الأيكة، ابتعد، القفرة، انتصر د) انتصر، القفرة، ابتعد، الأيكة

١

١١-(مَن مُخْتارُ هَذِهِ النصوصِ؟)، كلمة (مُختار) في الجملة السابقة هي:

أ) اسم فاعل ب) اسم مفعول ج) صيغة مبالغة د) صفة مشبّهة

١٢- الصواب في استخدام (اسم التفضيل) أن تقول:

أ) البنت الكبير ب) البنت الكبيرة ج) البنت الكبرى د) البنت الأكبر

١٣- " تَماثَلَ الطفْلُ لِلشفاءِ" المعنى المستفاد من الزيادة في الفعل المخطوط تحته:

أ) التعدية ب) المشاركة ج) الصيرورة د) التدريج

١٤- (تَحَدَّثَ إِلَيْنا الْخَطيبُ وَهُوَ واقِفٌ)، جملة (هو واقف) في محلّ:

أ) نصب حال ب) رفع صفة ج) جرّ مضاف إليه د) رفع خبر

١٥- الوزن الصرفي لكلمة (انْتَقَعَ)، هو:

أ) افْتَعَلَ ب) انْفَعَلَ ج) افْعَلَلَ د) افْعَلَّ

١٦- قال تعالى : " وقُلْ لَهُما قَوْلاً مَيْسورًا " ، الفعل (قُلْ) :

أ) معتلّ ناقص ب) صحيح سالم ج) معتلّ أجوف د) صحيح مضعّف

١٧- قال تعالى: "وَالذينَ إذا ذُكِّروا بِآياتِ رَبِّهِمْ لَمْ يَخِرّوا عَلَيْها صُمًّا وَعُمْيانا " الفعل (يَخِرّوا) يُعرب فعلاً مضارعًا:

أ) مرفوعًا وعلامة رفعه الواو ب) مجزومًا وعلامة جزمه السكون

ج) مجزومًا وعلامة جزمه حذف النون د) مرفوعًا وعلامة رفعه الألف

١٨- قال تعالى: " يا أَيُّها الْمُزَّمِّلُ قُمِ اللَيْلَ إلاّ قَليلا" إعراب كلمة (قليلا) هو:

أ) مستثنى منصوب ب) حال منصوب ج) تمييز منصوب د) مفعول فيه منصوب

١٩-إذا أردتُ التعجب من جمال السماء فإنني أقول:

أ) ما أجملَ السماءَ ب) ما أجملُ السماءِ ج) ما أجملَ السماءُ د) ما أجملَ شيءٍ في السماء

٢٠- وَلا أَحْمِلُ الْحِقْدَ الْقَديمَ عَلَيْهِمُ وَلَيْسَ رَئيسُ الْقَوْمِ مَنْ يَحْمِلُ الْحِقْدا

(لا) في البيت هي حرف:

أ) نهي ب) نفي ج) عطف د) جرّ

٢١- " رُبَّ لَحْدٍ قَدْ صارَ لَحْدًا مِرارًا ضاحِكٍ مِنْ تَزاحُمِ الأَضْدادِ " (قد) في البيت تفيد:

أ) التحقيق ب) التشكيك ج) التقليل د) التعليل

٢٢- (أَنْتَ .... مُجْتَهِدٌ). الكتابة الصحيحة للكلمة المناسبة لملء الفراغ هي:

أ) امرء ب) امرىء ج) امرؤ د) امروء

٢٣- **واحد ممّا تحته خطّ في ما يأتي كُتب خطأ:**

أ) اهتمّتِ الْحُكومَةُ بإنْشاءِ الْمَدارِسِ في الريفِ. ب) كَتَبْتُ مَوْضوعَ إِنْشاءٍ عَلى الْحاسوبِ.

ج) إِنْشاءَ اللهُ سَنَفوزُ في الْمُسابَقَةِ الشعْرِيَّةِ. د) هَذا الْجِسْرُ مِنْ إِنْشاءِ مُهَنْدِسينَ خُبَراءَ.

٢٤- **" أَقَمْتُ في مَكَّةَ عامَيْنِ ". حركة آخر كلمة (مكّة) هي:**

أ) الفتحة ب) الضمّة ج) الكسرة د) السكون

٢٥- **الجملة التي تُمثّل أسلوب الشرط من الجمل الآتية هي:**

أ) قال تعالى: " لَوْ أَنْزَلْنا هَذا الْقُرْآنَ عَلى جَبَلٍ لَرَأَيْتَهُ خاشِعًا مُتَصَدِّعًا مِنْ خَشْيَةِ اللهِ "

ب) قال تعالى: " قُلْ أعوذُ بِرَبِّ الناسِ "

ج) وَعاشِرْ بِمَعْروفٍ وَسامِحْ مَنِ اعْتَدى وَدافِعْ وَلَكِنْ بِالتي هِيَ أَحْسَنُ

د) قال تعالى: " لا يَسْخَرْ قَوْمٌ مِنْ قَوْمٍ عَسى أَنْ يَكونوا خَيْرًا مِنْهُمْ "

٢٦- **" وَجَيْشٍ كَجُنْحِ الليلِ يَزْحَفُ بِالْحَصى "، (الواو) في أول شطر البيت هي واو:**

أ) الاستئناف ب) ربّ ج) العطف د) المعيّة

٢٧- **قالَ خالِدٌ لِصَديقِهِ ... ما أَجْمَلَ مَدينَةَ الْبَتْراءِ ... هَلْ زُرْتَها...**

علامات الترقيم المناسبة لملء الفراغ في ما سبق على الترتيب هي:

أ) (: ! ؟) ب) (! : ؟) ج) (: ؟ ،) د) (. : !)

٢٨- **المقطع القصير في الكتابة العروضية هو حرف:**

أ) متحرّك ب) ساكن ج) متحرّك يليه ساكن د) ساكن يليه متحرّك

٢٩- **(وَطَني أَنْتَ الْحَبيبُ الدّائِمُ) الصورة الصحيحة لتقطيع هذا الشطر هي:**

أ) ∪ ∪ – – – ∪ – ∪ – – – ∪ – – – ∪ – ب) ∪ ∪ – – – ∪ – – – ∪ – – ∪ –

ج) ∪ ∪ – – – ∪ – – ∪ ∪ – د) ∪ ∪ – – – ∪ – – – – – – –

٣٠- **واحدة مما يأتي ليست من صور التفعيلة (مُسْتَفْعِلُنْ):**

أ) مُتَفْعِلُنْ ب) مُتَفاعِلُنْ ج) مُتَّعِلُنْ د) مُسْتَعِلُنْ

انتهت أسئلة اللغة العربية

تابع أسئلة الرياضيات من (٣١) إلى (٦٠)

٣١- يمكن إعادة تعريف الاقتران ق(س) = | س − ٧ | على النحو الآتي:

أ) ق(س) = { س − ٧ ، س ≤ ٧ ؛ ٧ − س ، س > ٧ }   ب) ق(س) = { ٧ − س ، س ≤ ٧ ؛ س − ٧ ، س > ٧ }

جـ) ق(س) = س − ٧   د) ق (س) = ٧ − س

٣٢- إذا كان ق(س) = { −س ، س ≤ ٣ ؛ ٢ ، س > ٣ } ، أوجد ق(٤) + ق(٣) ؟

أ) ٤   ب) −٢   جـ) −٧   د) −١

٣٣- إذا كان ق(س) = أس^{٢} + ١٧ ، هـ(س) = | ٢س + ١ | ، وكان (ق ο هـ) (−٢) = ٨ فإن قيمة أ هي:

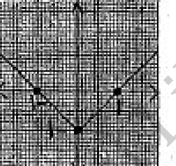
أ) −١   ب) $\frac{-٥}{٢}$ جـ) $\frac{-٢٧}{٨}$ د) $\frac{-٤٣}{٨}$

٣٤- حل المعادلة [$\frac{١}{٣}$ س + ١] = ٢ هو:



أ) [٢ ، ٣)   ب) [٣ ، ٦)   جـ) [١ ، ٢)   د) [٣ ، ∞)

٣٥- قاعدة الاقتران الممثل في الشكل المجاور هي:



أ) ق(س) = | س − ١ | ب) ق(س) = |١ − س |

جـ) ق(س) = | س | + ١   د) ق(س) = | س | − ١

٣٦- حل نظام المعادلات ٣س = ٦ ، ص + ٣ س = ٧ ، ٢س + ٣ص − ٢ع = ٥ هو:

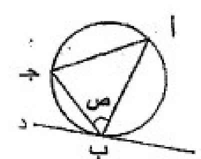
أ) س=٢ ، ص=−١ ، ع = −١٢   ب) س=٢ ، ص=١ ، ع =١

جـ) س=٢ ، ص=−١ ، ع = −٢   د) س=٢ ، ص=١ ، ع =−١

٣٧- ميل المستقيم ل$_{١}$ الموازي للمستقيم ل$_{٢}$ الذي يمر بالنقطتين (٢ ، ٥) ، (٤ ، ٩) يساوي :

أ) ١   ب) ٢   جـ) −١   د) −٢

٣٨- في الشكل المجاور، إذا كان قياس ∠جـ ب د = ٤٥° ، وقياس الزاوية ∠أ جـ ب = ٦٥° فإن قياس ∠ص بالدرجات يساوي:



أ) ٤٥°   ب) ١٢٠°   جـ) ٧٠°   د) ١١٠°

٤

٣٩- في الشكل المجاور دائرة مركزها م ، س ص مماس للدائرة في النقطة د، قياس ∠م ص د = ٤٠° ما قياس ∠د أ ب بالدرجات ؟

أ) ٢٥°　　ب) ٥٠°　　جـ) ٢٠°　　د) ٤٠°

٤٠- في الشكل المجاور أ ب ج د رباعي دائري، ما قيمة س بالدرجات ؟

أ) ٢٠°　　ب) ١٨°　　جـ) ١٥°　　د) ٣٦°

٤١- النقطة التي تمثل حلا لنظام المعادلات س^{٢} + ص^{٢} = ٥٠ ، س = ص هي:

أ) (-٥ ، ٥)　　ب) (-٥ ، √٥)　　جـ) (√٥ ، -√٥)　　د) (٥ ، ٥)

٤٢- إذا كانت س ، ص ، ع تمثل أعمار الابن والأب والأم على الترتيب في إحدى العائلات، وكان عمر الأب ضعف عمر الابن، وعمر الأم يساوي معدل عمري الأب والابن، ومجموع أعمار الأب والأم والابن ١٣٥ عاماً، فإن نظام المعادلات الذي يحل هذه المسألة هو:

أ) س + ص + ع = ١٣٥
٢س = ص
٢س + ص = ٢ع

ب) س + ص + ع = ١٣٥
س = ٢ص
س + ٢ص = ٢ع

جـ) س + ص + ع = ١٣٥
٢س = ص
س + ص = ٢ع

د) س + ص + ع = ١٣٥
س = ٢ص
س + ص = ٢ع

٤٣- رتبة المصفوفة $\begin{bmatrix} ٢ & ١ \\ ٣ & ١ \\ ٤ & ٥ \end{bmatrix}$ تساوي:

أ) ٢ × ٣　　ب) ٣ × ٢　　جـ) ٦ × ١　　د) ١ × ٦

٤٤- إذا كانت أ = $\begin{bmatrix} س - ٢ & ٢س \\ -١ & ٢ \end{bmatrix}$ مصفوفة منفردة فما قيمة س؟

أ) -١　　ب) ١　　جـ) صفر　　د) ٤

٤٥- التعبير الذي يمثل عمليات الصف البسيطة المطبقة على المصفوفة أ = $\begin{bmatrix} ٥ & -٦ \\ -١ & ٢ \end{bmatrix}$ حتى تصبح مساوية للمصفوفة ب = $\begin{bmatrix} ٢ & ٠ \\ -١ & ٢ \end{bmatrix}$ هو:

أ) ٣ص$_{١}$ + ص$_{٢}$ ⟵ ص$_{٢}$ ب) ٣ص$_{٢}$ + ص$_{١}$ ⟵ ص$_{٢}$

جـ) ٣ص$_{٢}$ + ص$_{١}$ ⟵ ص$_{١}$ د) ٣ص$_{١}$ + ص$_{٢}$ ⟵ ص$_{١}$

٤٦- إذا كانت المصفوفة س = $\begin{bmatrix} ٣ & -٢ & ١ \\ ١ & ٤ & ٥ \end{bmatrix} \times \begin{bmatrix} ٥ & ٤ \\ ٢ & -٣ \\ ٦ & ١ \end{bmatrix}$، فإن قيمة المدخلة س$_{٢٢}$ تساوي :

أ) ١٧ ب) ٣ جـ) ٢٣ د) -٣

٤٧- مجموعة حل نظام المعادلات س^{٢} + ص^{٢} = ٥ ، س^{٢} - ص^{٢} = ٣ هو:

أ) { (٢ ، ١) ، (-٢ ، -١) } ب) { (١ ، ٢) ، (-١ ، -٢) }

جـ) { (٢ ، ١) ، (-٢ ، -١) ، (٢ ، -١) ، (-٢ ، ١) } د) { (١ ، ٢) ، (-١ ، ٢) ، (١ ، -٢) ، (-١ ، -٢) }

٤٨- الشكل الذي يمثل الوضع القياسي للزاوية التي قياسها ٢٢٥° هو:

أ) ب) جـ) د)

٤٩- إذا كان جتا هـ = $\frac{٤}{٥}$ ، وكان ٢٧٠° < هـ < ٣٦٠° ، فإن جا هـ يساوي:

أ) - $\frac{٤}{٥}$ ب) - $\frac{٣}{٥}$ جـ) $\frac{٣}{٥}$ د) $\frac{٤}{٥}$

٥٠- جا^{٢} ٣١٥° + جتا^{٢} ١٣٥° يساوي :

أ) ١ ب) -١ جـ) صفر د) -$\sqrt{٢}$

٥١- ما عدد الزوايا التي تحقق العلاقة جاهـ = $\sqrt{٣}$ جتاهـ ، حيث ٠° < هـ < ٣٦٠° ؟

أ) زاوية واحدة ب) ثلاث زوايا جـ) زاويتان د) أربع زوايا

٥٢- في الشكل المجاور إذا كانت د منتصف $\overline{أجـ}$ ، فإن طول $\overline{ب د}$ يساوي:

أ ، د ، ٦ ، ب ، جـ ، ٨

أ) ١ ب) ٥ جـ) ١٠ د) ٢٠

٥٣- إذا كان الوسط الحسابي لعلامات طالب في أربعة اختبارات هو ٧٠ ، وراجع الطالب المعلم في أحد الاختبارات وكانت علامته ٨٠ وأصبحت ٨٨، كم يصبح الوسط الحسابي بعد المراجعة :

أ) ٧٢ ب) ٨٢ جـ) ٧٨ د) ٦٨

٥٤- قيمة المدى للتوزيع التكراري الموضح في الجدول المجاور هي:

أ) ٤ ب) ٥ جـ) ١٩ د) ٢٠

الفئات	٢٠-٢٤	٢٥-٢٩	٣٠-٣٤	٣٥-٣٩
التكرار	٦	٦	٧	١٠

٥٥- أي المنحنيات التكرارية المرقمة في الشكل المجاور يمثل التوزيع الأقل تشتتا:

أ) ٤   ب) ٣

جـ) ٢   د) ١

٥٦- إذا كان الانحراف المعياري لمجموعة مشاهدات يساوي (٤)، وعدلت جميع المشاهدات حسب العلاقة (ص=-٢س+٥)، حيث س هي المشاهدات قبل التعديل، ص المشاهدات بعد التعديل، فإن قيمة الانحراف بعد التعديل تساوي:

أ) -١٣   ب) ١٣   جـ) ٨   د) -٨

٥٧- إذا كان $ح_١$، $ح_٢$ حادثان منفصلان في تجربة عشوائية، وكان ل ($ح_١$) = ٠,٤ ، ل ($ح_٢$) =٠,٥
فجد ل ($ح_١$ − $ح_٢$).

أ) صفر   ب) ٠,٤   جـ) ٠,٥   د) ٠,١

٥٨- في تجربة إلقاء حجري نرد منتظمين مرة واحدة، وتسجيل عدد النقاط الظاهرة على الوجه العلوي لكل منهما، ما احتمال الحصول على عددين مجموعهما يقبل القسمة على (٦) ؟

أ) $\frac{١}{٦}$ ب) $\frac{١}{٤}$ جـ) $\frac{١}{٣}$ د) $\frac{١}{٢}$

٥٩- في تجربة عشوائية فضاؤها العيني Ω ، إذا كان $ح_١ \subset ح_٢$، ل($ح_٢$)= ٠,٨ ، ل($ح_٢$ − $ح_١$)= ٠,٥ ، فإن ل($ح_١ \cap ح_٢$) يساوي:

أ) ٠,٢٤   ب) ٠,٨   جـ) ٠,٥   د) ٠,٣

٦٠- إذا كان احتمال نجاح زراعة كلية للمريض الأول (٠,٤)، واحتمال نجاح زراعة كلية للمريض الثاني (٠,٥)، فإن احتمال نجاح زراعة الكلية لأحدهما على الأقل يساوي:

أ) ٠,٤   ب) ٠,٥   جـ) ٠,٧   د) ٠,٩

انتهت الأسئلة بحمد الله

ملحق (3)

الاختبار الوطني

مفتاح الإجابة

الصف : العاشر

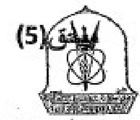
المبحث : العلوم

الرقم	رمز الإجابة الصحيحة	المحور
٣١	ب	الأحياء
٣٢	د	
٣٣	أ	
٣٤	ب	
٣٥	د	
٣٦	د	
٣٧	ج	
٣٨	ج	
٣٩	د	الفيزياء
٤٠	ب	
٤١	ب	
٤٢	د	
٤٣	ب	
٤٤	أ	
٤٥	ج	
٤٦	أ	
٤٧	ج	
٤٨	ب	الكيمياء
٤٩	أ	
٥٠	ج	
٥١	ب	
٥٢	ج	
٥٣	د	
٥٤	د	
٥٥	ب	
٥٦	د	علوم الأرض
٥٧	ب	
٥٨	ب	
٥٩	ج	
٦٠	د	

الملحق (4)

ملحق (5)

YARMOUK UNIVERSITY
Office of the President



جامعة اليرموك
مكتب الرئيس

Date: ....

Reference: ....

التاريخ : ....
الموافق : ١٦/ ربيع الآخر/ ١٤٣٤م
الرقم : ٦/ ٢٦/ شباط/ ٢٠١٣

معالي وزير التربية والتعليم والتعليم العالي والبحث العلمي الأكرم

الموضوع: تسهيل مهمة الطالبة نجود علي محمود فريحات

تحية طيبة وبعد ،،،

تقوم الطالبة نجود علي محمود فريحات، ورقمها الجامعي (٢٠١٠٢٢٠٠٠٦)، بدراسة بعنوان " الأداء التفاضلي على الاختبار الوطني لضبط نوعية التعليم للصف العاشر من حيث شكل الاختبار ومحتوى المادة"؛ وذلك استكمالاً لمتطلبات الحصول على درجة الدكتوراة في كلية التربية، تخصص قياس وتقويم، ويستدعي ذلك الحصول على بيانات تتعلق بأداة الدراسة (اختبار الوطني لضبط نوعية التعليم للصف العاشر) على طلبة الصف العاشر في المملكة.

أرجو التكرم بالاطلاع والموافقة على تسهيل مهمة الطالبة المذكورة أعلاه.

شاكراً لكم تعاونكم مع الجامعة.

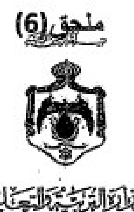
وتفضلوا بقبول فائق الاحترام،،،

رئيس الجامعة

أ.د. عبدالله الموسى

إربد - الأردن
الهاتف: ٧٢١١١١١ - ٢ - ٩٦٢ فاكس: ٧٢١١١٣٣ - ٢ - ٩٦٢
www.yu.edu.jo P.O. Box 566, Irbid, Jordan P: +962 2 7211111 F: +962 2 7211133 E-mail: yarmouk@yu.edu.jo president@yu.edu.jo

ملحق (6)



وزارة التربية والتعليم

الرقم ٣/١٠/ ٩٥٣٤
التاريخ ٢٤ ربيع الثاني ١٤٣٤
الموافق ٢٠١٣/٠٣/٠٧

السيد مدير إدارة الامتحانات والاختبارات

الموضوع: البحث التربوي

السلام عليكم ورحمة الله وبركاته

تقوم الطالبة نجود علي محمود فريحات بإجراء دراسة عنوانها " الأداء التفاضلي على الاختبار الوطني لضبط نوعية التعليم للصف العاشر من حيث شكل الاختبار ومحتوى المادة "، وذلك استكمالا لمتطلبات الحصول على درجة الدكتوراه تخصص قياس وتقويم في جامعة اليرموك، ويحتاج ذلك إلى معلومات وبيانات إحصائية من إدارتكم.

يرجى تسهيل مهمة الطالبة المذكورة وتقديم المساعدة الممكنة لها.

وتفضلوا بقبول فائق الاحترام

وزير التربية والتعليم

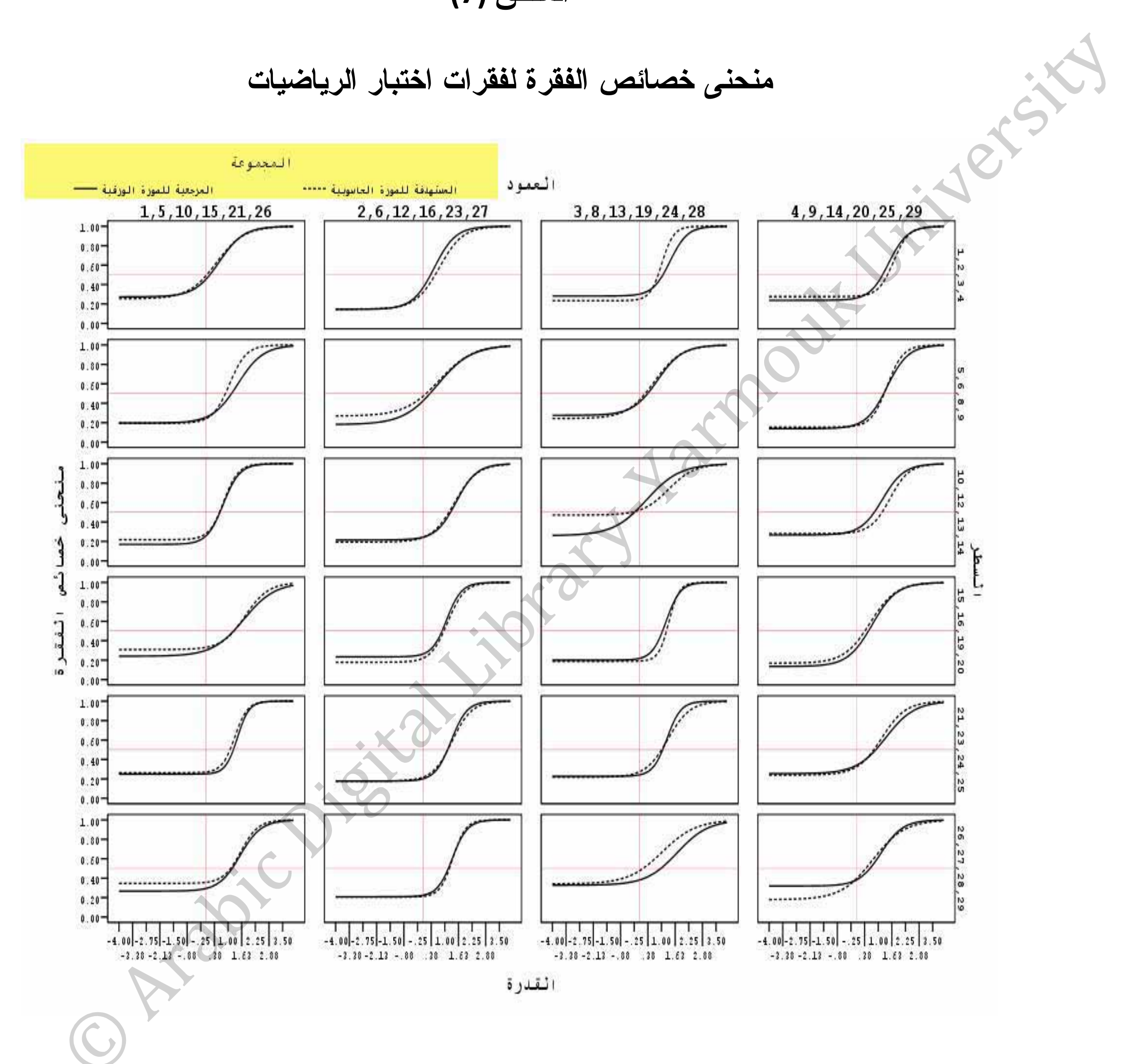
الدكتور
مهند خليل المباسنة
مدير البحث والتطوير التربوي

نسخة/ رئيس قسم البحث التربوي
نسخة/ الملف ١٠/٢

المملكة الأردنية الهاشمية
هاتف ٥٦٠٧١٨١ ٦ ٩٦٢+ فاكس ٥٦٦٦٠١٩ ٦ ٩٦٢+ ص.ب ١٦٤٦ عمان ١١١١٨ الأردن الموقع الإلكتروني www.moe.gov.jo

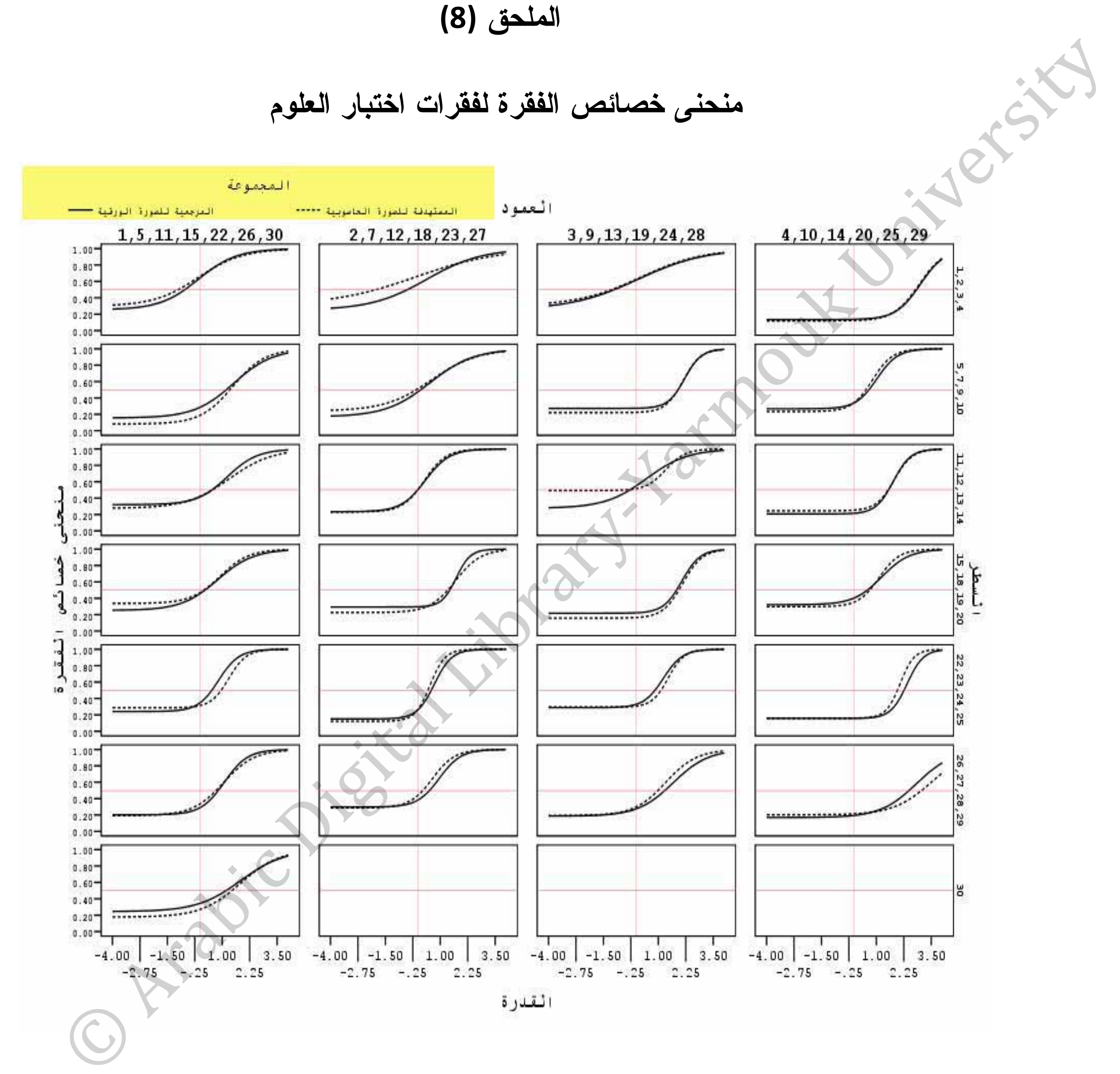
الملحق (7)

منحنى خصائص الفقرة لفقرات اختبار الرياضيات



الملحق (8)

منحنى خصائص الفقرة لفقرات اختبار العلوم



Abstract

Friehat, Nujood Ali Mohmoud. (2013) "Differential item functioning of the national educational quality control test in mathematics and science for 10th grade according to mode of test presentation" PhD dissertation, Yarmouk University. (supervisor: Prof. Dr. Sari Salim Sawaqed).

This study aimed at investigating the differential functioning of science and mathematics test and their items resulting from education quality control test in accordance to mode of test presentation (paper or computerized)

To achieve the purpose of the study the data of a sample consisted of 2764 students from those who took the test in the year 2010/2011 were used. the final numbers for science and mathematics items were (25, 24) of the multiple choice type where every item had four alternatives, the science test items measure themes related to physics, chemistry, earth science and biology, while the items of mathematics test measure algebra, geometry and measurement, statistics and probability.

The items of the test were assessed in accordance to the three parameter model where the differential functioning of the items was revealed using the method of "the differential functioning of item and tests (DIFT)" where the (NCDIF) was used as a marker for differential functioning for the items and the (CDIF) as a marker for the differential functioning for the items relevance to the test. The statistical programs

(SPSS, BILOG−MG 3, Equate v2.1, Dfit v8.04) were used in the analysis process and the most prominent results were:

- There is differential item functioning at the mathematics test in relevance to the test presentation with a percentage reached (8,33%)with two items out from the sum of all items, one which showed uniform differential item functioning for the benefit of computerized test form while the other showed non uniform differential item functioning for both test forms.
- There is differential item functioning at the science test in relevance to the test presentation with a percentage reached (12%) with three items all of which showed non uniform differential item functioning sometimes for the paper form of test and sometimes for computerized one.
- The study attributed the reasons for the emergence of differential functioning in items of science and mathematics tests for the benefit of computerized test to the nature of items content which requires mental skills where special notations on side paper sheets are not required to reach the correct answer as there was a smoothness in dealing with test items where tested student benefited from the advantages of computerized tests especially those related to the factors, of movement like, the ability of deletion of wrong answers as it was faster in computerized form compared to the traditional one or that is related to how to respond to the item by shading the correct alternative with a pencil on paper form and clicking the

mouse on the computerized form which increased the factor of concentration of the tested student and kept him away from guessing to which he resorted on the paper form. On the other hand, the items that require special notation on side paper sheets increased the factor of confusion on the computerized form and hindered the extent of his benefit from the advantages of this test and so the paper form became a suitable choice that keeps the tested student away from guessing that he resorted to on computerized form.

- Most of the items of the mathematics and science Test were free of differential functioning as 22 items of the two tests did not show differential item functioning in relation to the variable depending on the form of the test, by approximately 91.67% to the items of mathematics test, and 88% to the items of science test, and thus may be possible to rely on these items in the future for the development of tests free of differential functioning to a reasonable extent.
- The items that showed differential functioning at the level of item were few. Thus, practically there will be no great injustice related to form of test between tested students, a result of great importance for test developers that can be used accurately to assess the extent of the possession of tenth grade students for basic skills of learning in the areas of science, mathematics and helps to make important decisions about the reality of education in Jordan.

- In general the test as a whole did not show significant statistical differential functioning in science and mathematics tests depending on the test presentation. Thus, there may not be a problem in the transition to computerized tests as a reliable just alternative to the paper form, which may lead to the expansion in the development of computerized tests.

In light of this, the study recommends that further studies should be performed to investigate differential item functioning of the national quality control test to be more comprehensive to cover the four areas tested (science, mathematics, Arabic language, English Language) and to cover the fourth , eighth and tenth grades to compare between them according to the item response theory.

Keywords:

Differential item function, differential functioning of item and test, item response theory, science and mathematics tests taken from national education quality test, three parameter model